

CE

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

**(10) International Publication Number**  
**WO 00/78960 A2**

- [Continued on next page]*

```

1  HFFPQLQGVYINLYTPFRFVAAAGUGGUSSAGERSQFUAHASELAAGSG  N1002Cof1.FRO
   Ireqoqls.FRO
1  LSNVLSESVYCAPFAAAAAAAAAAAAGYGATLPYATLELPYFVQLGAUYELK  N1002Cof1.FRO
51 Ireqoqls.FRO
1  GPGVQRTATAAAFFPHFPAFYPTQVYQFQFHPFRFAHATRELSSTLRAMW  N1002Cof1.FRO
101 Ireqoqls.FRO
1  PPPRRPFYPTKQEKINLAITENTMLTQVSTWFAHARRIKKKKHTMAFAS  N1002Cof1.FRO
151 Ireqoqls.FRO
1  YDERQGNATQREKEEDESEDFEERERELEKFKERPLAGRCHHHNGGUCAC  N1002Cof1.FRO
201 Ireqoqls.FRO
1  KRRRDEKIDLENLDSAAAGSELYLAGAAHHHHGFLGPISDCKTSSDSUSS  N1002Cof1.FRO
   Ireqoqls.FRO
11  FERRDAGGQFSGGOREGQARRRFGQFAPRRHHVHVFSLHFFAFAPAFAPFALOK  N1002Cof1.FRO
301 Ireqoqls.FRO
1  KIMSLAKYATQDRPHRRRDFQAGGFFFGAAVAFTNLSFAAAAAAAAF  N1002Cof1.FRO
351 Ireqoqls.FRO
1  KIMSLASTATYDRPHRRRDFQAGGFFFGAAVAFTNLSFAAAAAAAAF  N1002Cof1.FRO
104  FLVSAFLQGFPAWTHRRFFQFQFQFPHFPLSLHFGSAFPHHNLFGFAAHN  N1002Cof1.FRO
401 Ireqoqls.FRO
1  AAAAALFAPFAFFGQFQFRCALFVKKKKLYTAFQFVFVFQFHHNLDAAL  N1002Cof1.FRO
451 Ireqoqls.FRO
208  LRALRSN  N1002Cof1.FRO
508  LVALRSN  Ireqoqls.FRO

```

**(57) Abstract:** Compositions and methods for the therapy and diagnosis of cancer, such as breast cancer, are disclosed. Compositions may comprise one or more breast tumor proteins, immunogenic portions thereof, or polynucleotides that encode such portions. Alternatively, a therapeutic composition may comprise an antigen presenting cell that expresses a breast tumor protein, or a T cell that is specific for cells expressing such a protein. Such compositions may be used, for example, for the prevention and treatment of diseases such as breast cancer. Diagnostic methods based on detecting a breast tumor protein, or mRNA encoding such a protein, in a sample are also provided.

**WO 00/78960 A2**

**THIS PAGE BLANK (USPTO)**



patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

**Published:**

— *Without international search report and to be republished upon receipt of that report.*

## COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS OF BREAST CANCER

### TECHNICAL FIELD OF THE INVENTION

The present invention relates generally to therapy and diagnosis of cancer, such as breast cancer. The invention is more specifically related to polypeptides comprising at least a portion of a breast tumor protein, and to polynucleotides encoding such polypeptides. Such polypeptides and polynucleotides may be used in vaccines and pharmaceutical compositions for prevention and treatment of breast cancer, and for the diagnosis and monitoring of such cancers.

### 10 BACKGROUND OF THE INVENTION

Breast cancer is a significant health problem for women in the United States and throughout the world. Although advances have been made in detection and treatment of the disease, breast cancer remains the second leading cause of cancer-related deaths in women, affecting more than 180,000 women in the United States each year. For women in North America, the life-time odds of getting breast cancer are now one in eight.

No vaccine or other universally successful method for the prevention or treatment of breast cancer is currently available. Management of the disease currently relies on a combination of early diagnosis (through routine breast screening procedures) and aggressive treatment, which may include one or more of a variety of treatments such as surgery, radiotherapy, chemotherapy and hormone therapy. The course of treatment for a particular breast cancer is often selected based on a variety of prognostic parameters, including an analysis of specific tumor markers. *See, e.g., Porter-Jordan and Lippman, Breast Cancer 8:73-100, 1994.* However, the use of established markers often leads to a result that is difficult to interpret, and the high mortality observed in breast cancer patients indicates that improvements are needed in the treatment, diagnosis and prevention of the disease.

Immunotherapies have the potential to substantially improve breast cancer treatment and survival. Such therapies may involve the generation or

enhancement of an immune response to a breast tumor antigen. However, to date, relatively few breast tumor antigens are known and the generation of an immune response against such antigens has not been shown to be therapeutically beneficial.

Accordingly, there is a need in the art for improved methods for  
5 identifying breast tumor antigens and for using such antigens in the diagnosis and therapy of breast cancer. The present invention fulfills these needs and further provides other related advantages.

#### SUMMARY OF THE INVENTION

Briefly stated, the present invention provides compositions and methods  
10 for the diagnosis and therapy of cancer, such as breast cancer. In one aspect, the present invention provides polypeptides comprising at least a portion of a breast tumor protein, or a variant thereof. Certain portions and other variants are immunogenic, such that the ability of the variant to react with antigen-specific antisera is not substantially diminished. Within certain embodiments, the polypeptide comprises a sequence that is  
15 encoded by a polynucleotide sequence selected from the group consisting of sequences recited in SEQ ID NOs:1-125, variants of such sequences and complements of such sequences. One such polypeptide comprises a sequence recited in SEQ ID NO:126, or a variant thereof that is at least 90% identical to SEQ ID NO:126.

The present invention further provides polynucleotides that encode a  
20 polypeptide as described above, or a portion thereof (such as a portion encoding at least 9, preferably at least 15, amino acid residues of a breast tumor protein), expression vectors comprising such polynucleotides and host cells transformed or transfected with such expression vectors.

Within other aspects, the present invention provides pharmaceutical  
25 compositions comprising a polypeptide or polynucleotide as described above and a physiologically acceptable carrier.

Within a related aspect of the present invention, vaccines are provided. Such vaccines comprise a polypeptide or polynucleotide as described above and a non-specific immune response enhancer.

The present invention further provides pharmaceutical compositions that comprise: (a) an antibody or antigen-binding fragment thereof that specifically binds to a breast tumor protein; and (b) a physiologically acceptable carrier.

Within further aspects, the present invention provides pharmaceutical  
5 compositions comprising: (a) an antigen presenting cell that expresses a polypeptide as described above and (b) a pharmaceutically acceptable carrier or excipient. Antigen presenting cells include dendritic cells, macrophages and B cells.

Within related aspects, vaccines are provided that comprise: (a) an  
10 antigen presenting cell that expresses a polypeptide as described above and (b) a non-specific immune response enhancer.

The present invention further provides, in other aspects, fusion proteins that comprise at least one polypeptide as described above, as well as polynucleotides encoding such fusion proteins.

Within related aspects, pharmaceutical compositions comprising a fusion  
15 protein, or a polynucleotide encoding a fusion protein, in combination with a physiologically acceptable carrier are provided.

Vaccines are further provided, within other aspects, that comprise a fusion protein or a polynucleotide encoding a fusion protein in combination with a non-specific immune response enhancer.

20 Within further aspects, the present invention provides methods for inhibiting the development of a cancer in a patient, comprising administering to a patient a pharmaceutical composition or vaccine as recited above.

The present invention further provides, within other aspects, methods for removing tumor cells from a biological sample, comprising contacting a biological  
25 sample with T cells that specifically react with a breast tumor protein, wherein the step of contacting is performed under conditions and for a time sufficient to permit the removal of cells expressing the protein from the sample.

Within related aspects, methods are provided for inhibiting the development of a cancer in a patient, comprising administering to a patient a biological  
30 sample treated as described above.

Methods are further provided, within other aspects, for stimulating and/or expanding T cells specific for a breast tumor protein, comprising contacting T cells with one or more of: (i) a polypeptide as described above; (ii) a polynucleotide encoding such a polypeptide; and/or (iii) an antigen presenting cell that expresses such a polypeptide; under conditions and for a time sufficient to permit the stimulation and/or expansion of T cells. Isolated T cell populations comprising T cells prepared as described above are also provided.

Within further aspects, the present invention provides methods for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of a T cell population as described above.

The present invention further provides methods for inhibiting the development of a cancer in a patient, comprising the steps of: (a) incubating CD4<sup>+</sup> and/or CD8<sup>+</sup> T cells isolated from a patient with one or more of: (i) a polypeptide comprising at least an immunogenic portion of a breast tumor protein; (ii) a polynucleotide encoding such a polypeptide; and (iii) an antigen-presenting cell that expresses such a polypeptide; and (b) administering to the patient an effective amount of the proliferated T cells, and thereby inhibiting the development of a cancer in the patient. Proliferated cells may, but need not, be cloned prior to administration to the patient.

Within further aspects, the present invention provides methods for determining the presence or absence of a cancer in a patient, comprising (a) contacting a biological sample obtained from a patient with a binding agent that binds to a polypeptide as recited above; (b) detecting in the sample an amount of polypeptide that binds to the binding agent; and (c) comparing the amount of polypeptide with a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient. Within preferred embodiments, the binding agent is an antibody, more preferably a monoclonal antibody. The cancer may be breast cancer.

The present invention also provides, within other aspects, methods for monitoring the progression of a cancer in a patient. Such methods comprise the steps of: (a) contacting a biological sample obtained from a patient at a first point in time with a binding agent that binds to a polypeptide as recited above; (b) detecting in the

sample an amount of polypeptide that binds to the binding agent; (c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and (d) comparing the amount of polypeptide detected in step (c) with the amount detected in step (b) and therefrom monitoring the progression of the cancer in the  
5 patient.

The present invention further provides, within other aspects, methods for determining the presence or absence of a cancer in a patient, comprising the steps of: (a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a breast tumor protein; (b) detecting in the  
10 sample a level of a polynucleotide, preferably mRNA, that hybridizes to the oligonucleotide; and (c) comparing the level of polynucleotide that hybridizes to the oligonucleotide with a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient. Within certain embodiments, the amount of mRNA is detected via polymerase chain reaction using, for example, at least one  
15 oligonucleotide primer that hybridizes to a polynucleotide encoding a polypeptide as recited above, or a complement of such a polynucleotide. Within other embodiments, the amount of mRNA is detected using a hybridization technique, employing an oligonucleotide probe that hybridizes to a polynucleotide that encodes a polypeptide as recited above, or a complement of such a polynucleotide.

20 In related aspects, methods are provided for monitoring the progression of a cancer in a patient, comprising the steps of: (a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a breast tumor protein; (b) detecting in the sample an amount of a polynucleotide that hybridizes to the oligonucleotide; (c) repeating steps (a) and (b)  
25 using a biological sample obtained from the patient at a subsequent point in time; and (d) comparing the amount of polynucleotide detected in step (c) with the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

Within further aspects, the present invention provides antibodies, such as  
30 monoclonal antibodies, that bind to a polypeptide as described above, as well as



diagnostic kits comprising such antibodies. Diagnostic kits comprising one or more oligonucleotide probes or primers as described above are also provided.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

#### BRIEF DESCRIPTION OF THE DRAWING

Figure 1 represents a sequence alignment between the representative breast tumor polypeptide B1002C (SEQ ID NO:126) and the mouse iroquois homeobox protein 3.

#### DETAILED DESCRIPTION OF THE INVENTION

As noted above, the present invention is generally directed to compositions and methods for using the compositions, for example in the therapy and diagnosis of cancer, such as breast cancer. Certain illustrative compositions described herein include breast tumor polypeptides, polynucleotides encoding such polypeptides, binding agents such as antibodies, antigen presenting cells (APCs) and/or immune system cells (*e.g.*, T cells). A "breast tumor protein," as the term is used herein, refers generally to a protein that is expressed in breast tumor cells at a level that is at least two fold, and preferably at least five fold, greater than the level of expression in a normal tissue, as determined using a representative assay provided herein. Certain breast tumor proteins are tumor proteins that react detectably (within an immunoassay, such as an ELISA or Western blot) with antisera of a patient afflicted with breast cancer.

Therefore, in accordance with the above, and as described further below, the present invention provides illustrative polynucleotide compositions having sequences set forth in SEQ ID NO:1-125, illustrative polypeptide compositions having amino acid sequences set forth in SEQ ID NO:126, antibody compositions capable of binding such polypeptides, and numerous additional embodiments employing such compositions, for example in the detection, diagnosis and/or therapy of human breast cancer.

**POLYNUCLEOTIDE COMPOSITIONS**

As used herein, the terms "DNA segment" and "polynucleotide" refer to a DNA molecule that has been isolated free of total genomic DNA of a particular species. Therefore, a DNA segment encoding a polypeptide refers to a DNA segment  
5 that contains one or more coding sequences yet is substantially isolated away from, or purified free from, total genomic DNA of the species from which the DNA segment is obtained. Included within the terms "DNA segment" and "polynucleotide" are DNA segments and smaller fragments of such segments, and also recombinant vectors, including, for example, plasmids, cosmids, phagemids, phage, viruses, and the like.

10 As will be understood by those skilled in the art, the DNA segments of this invention can include genomic sequences, extra-genomic and plasmid-encoded sequences and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides, peptides and the like. Such segments may be naturally isolated, or modified synthetically by the hand of man.

15 "Isolated," as used herein, means that a polynucleotide is substantially away from other coding sequences, and that the DNA segment does not contain large portions of unrelated coding DNA, such as large chromosomal fragments or other functional genes or polypeptide coding regions. Of course, this refers to the DNA segment as originally isolated, and does not exclude genes or coding regions later added  
20 to the segment by the hand of man.

As will be recognized by the skilled artisan, polynucleotides may be single-stranded (coding or antisense) or double-stranded, and may be DNA (genomic, cDNA or synthetic) or RNA molecules. RNA molecules include HnRNA molecules, which contain introns and correspond to a DNA molecule in a one-to-one manner, and  
25 mRNA molecules, which do not contain introns. Additional coding or non-coding sequences may, but need not, be present within a polynucleotide of the present invention, and a polynucleotide may, but need not, be linked to other molecules and/or support materials.

Polynucleotides may comprise a native sequence (*i.e.*, an endogenous  
30 sequence that encodes a breast tumor protein or a portion thereof) or may comprise a variant, or a biological or antigenic functional equivalent of such a sequence.

Polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions, as further described below, preferably such that the immunogenicity of the encoded polypeptide is not diminished, relative to a native tumor protein. The effect on the immunogenicity of the encoded polypeptide may generally be assessed as described herein. The term "variants" also encompasses homologous genes of xenogenic origin.

When comparing polynucleotide or polypeptide sequences, two sequences are said to be "identical" if the sequence of nucleotides or amino acids in the two sequences is the same when aligned for maximum correspondence, as described below. Comparisons between two sequences are typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A "comparison window" as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, M.O. (1978) A model of evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff, M.O. (ed.) *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenies pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) *CABIOS* 5:151-153; Myers, E.W. and Muller W. (1988) *CABIOS* 4:11-17; Robinson, E.D. (1971) *Comb. Theor* 11:105; Santou, N. Nes, M. (1987) *Mol. Biol. Evol.* 4:406-425; Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy – the Principles and Practice of Numerical Taxonomy*, Freeman Press, San Francisco, CA; Wilbur, W.J. and Lipman, D.J. (1983) *Proc. Natl. Acad., Sci. USA* 80:726-730.

Alternatively, optimal alignment of sequences for comparison may be conducted by the local identity algorithm of Smith and Waterman (1981) *Add. APL*.

*Math* 2:482, by the identity alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity methods of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics  
5 Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

One preferred example of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.* (1977) *Nucl. Acids Res.* 25:3389-3402  
10 and Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410, respectively. BLAST and BLAST 2.0 can be used, for example with the parameters described herein, to determine percent sequence identity for the polynucleotides and polypeptides of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. In one illustrative example, cumulative scores can be  
15 calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix can be used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved  
20 value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, and expectation (E) of 10, and the BLOSUM62 scoring matrix  
25 (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments, (B) of 50, expectation (E) of 10, M=5, N=-4 and a comparison of both strands.

Preferably, the "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a window of comparison of at least 20 positions, wherein the portion of the polynucleotide or polypeptide sequence in the  
30 comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less, usually 5 to 15 percent, or 10 to 12 percent, as compared to the reference

sequences (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid bases or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the reference sequence (*i.e.*, the window size) and multiplying the results by 100 to yield the percentage of sequence identity.

Therefore, the present invention encompasses polynucleotide and polypeptide sequences having substantial identity to the sequences disclosed herein, for example those comprising at least 50% sequence identity, preferably at least 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% or higher, sequence identity compared to a polynucleotide or polypeptide sequence of this invention using the methods described herein, (*e.g.*, BLAST analysis using standard parameters, as described below). One skilled in this art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like.

In additional embodiments, the present invention provides isolated polynucleotides and polypeptides comprising various lengths of contiguous stretches of sequence identical to or complementary to one or more of the sequences disclosed herein. For example, polynucleotides are provided by this invention that comprise at least about 15, 20, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500 or 1000 or more contiguous nucleotides of one or more of the sequences disclosed herein as well as all intermediate lengths there between. It will be readily understood that "intermediate lengths", in this context, means any length between the quoted values, such as 16, 17, 18, 19, *etc.*; 21, 22, 23, *etc.*; 30, 31, 32, *etc.*; 50, 51, 52, 53, *etc.*; 100, 101, 102, 103, *etc.*; 150, 151, 152, 153, *etc.*; including all integers through 200-500; 500-1,000, and the like.

The polynucleotides of the present invention, or fragments thereof, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their





overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed; with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol. For example, illustrative DNA segments with total lengths of about 10,000, about 5000,  
5 about 3000, about 2,000, about 1,000, about 500, about 200, about 100, about 50 base pairs in length, and the like, (including all intermediate lengths) are contemplated to be useful in many implementations of this invention.

In other embodiments, the present invention is directed to polynucleotides that are capable of hybridizing under moderately stringent conditions to  
10 a polynucleotide sequence provided herein, or a fragment thereof, or a complementary sequence thereof. Hybridization techniques are well known in the art of molecular biology. For purposes of illustration, suitable moderately stringent conditions for testing the hybridization of a polynucleotide of this invention with other polynucleotides include prewashing in a solution of 5 X SSC, 0.5% SDS, 1.0 mM  
15 EDTA (pH 8.0); hybridizing at 50°C-65°C, 5 X SSC, overnight; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS.

Moreover, it will be appreciated by those of ordinary skill in the art that, as a result of the degeneracy of the genetic code, there are many nucleotide sequences  
20 that encode a polypeptide as described herein. Some of these polynucleotides bear minimal homology to the nucleotide sequence of any native gene. Nonetheless, polynucleotides that vary due to differences in codon usage are specifically contemplated by the present invention. Further, alleles of the genes comprising the polynucleotide sequences provided herein are within the scope of the present invention.  
25 Alleles are endogenous genes that are altered as a result of one or more mutations, such as deletions, additions and/or substitutions of nucleotides. The resulting mRNA and protein may, but need not, have an altered structure or function. Alleles may be identified using standard techniques (such as hybridization, amplification and/or database sequence comparison).



**PROBES AND PRIMERS**

In other embodiments of the present invention, the polynucleotide sequences provided herein can be advantageously used as probes or primers for nucleic acid hybridization. As such, it is contemplated that nucleic acid segments that comprise  
5 a sequence region of at least about 15 nucleotide long contiguous sequence that has the same sequence as, or is complementary to, a 15 nucleotide long contiguous sequence disclosed herein will find particular utility. Longer contiguous identical or complementary sequences, *e.g.*, those of about 20, 30, 40, 50, 100, 200, 500, 1000 (including all intermediate lengths) and even up to full length sequences will also be of  
10 use in certain embodiments.

The ability of such nucleic acid probes to specifically hybridize to a sequence of interest will enable them to be of use in detecting the presence of complementary sequences in a given sample. However, other uses are also envisioned, such as the use of the sequence information for the preparation of mutant species  
15 primers, or primers for use in preparing other genetic constructions.

Polynucleotide molecules having sequence regions consisting of contiguous nucleotide stretches of 10-14, 15-20, 30, 50, or even of 100-200 nucleotides or so (including intermediate lengths as well), identical or complementary to a polynucleotide sequence disclosed herein, are particularly contemplated as  
20 hybridization probes for use in, *e.g.*, Southern and Northern blotting. This would allow a gene product, or fragment thereof, to be analyzed, both in diverse cell types and also in various bacterial cells. The total size of fragment, as well as the size of the complementary stretch(es), will ultimately depend on the intended use or application of the particular nucleic acid segment. Smaller fragments will generally find use in  
25 hybridization embodiments, wherein the length of the contiguous complementary region may be varied, such as between about 15 and about 100 nucleotides, but larger contiguous complementarity stretches may be used, according to the length complementary sequences one wishes to detect.

The use of a hybridization probe of about 15-25 nucleotides in length  
30 allows the formation of a duplex molecule that is both stable and selective. Molecules having contiguous complementary sequences over stretches greater than 15 bases in

length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having gene-complementary stretches of 15 to 25 contiguous nucleotides, or even longer where  
5 desired.

Hybridization probes may be selected from any portion of any of the sequences disclosed herein. All that is required is to review the sequence set forth in SEQ ID NO:1-125, or to any continuous portion of the sequence, from about 15-25 nucleotides in length up to and including the full length sequence, that one wishes to  
10 utilize as a probe or primer. The choice of probe and primer sequences may be governed by various factors. For example, one may wish to employ primers from towards the termini of the total sequence.

Small polynucleotide segments or fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, as is commonly  
15 practiced using an automated oligonucleotide synthesizer. Also, fragments may be obtained by application of nucleic acid reproduction technology, such as the PCR<sup>TM</sup> technology of U. S. Patent 4,683,202 (incorporated herein by reference), by introducing selected sequences into recombinant vectors for recombinant production, and by other recombinant DNA techniques generally known to those of skill in the art of molecular  
20 biology.

The nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of the entire gene or gene fragments of interest. Depending on the application envisioned, one will typically desire to employ varying conditions of hybridization to achieve varying degrees of  
25 selectivity of probe towards target sequence. For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, *e.g.*, one will select relatively low salt and/or high temperature conditions, such as provided by a salt concentration of from about 0.02 M to about 0.15 M salt at temperatures of from about 50°C to about 70°C. Such selective conditions tolerate  
30 little, if any, mismatch between the probe and the template or target strand, and would be particularly suitable for isolating related sequences.

Of course, for some applications, for example, where one desires to prepare mutants employing a mutant primer strand hybridized to an underlying template, less stringent (reduced stringency) hybridization conditions will typically be needed in order to allow formation of the heteroduplex. In these circumstances, one  
5 may desire to employ salt conditions such as those of from about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Cross-hybridizing species can thereby be readily identified as positively hybridizing signals with respect to control hybridizations. In any case, it is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide, which serves to  
10 destabilize the hybrid duplex in the same manner as increased temperature. Thus, hybridization conditions can be readily manipulated, and thus will generally be a method of choice depending on the desired results.

#### POLYNUCLEOTIDE IDENTIFICATION AND CHARACTERIZATION

Polynucleotides may be identified, prepared and/or manipulated using  
15 any of a variety of well established techniques. For example, a polynucleotide may be identified, as described in more detail below, by screening a microarray of cDNAs for tumor-associated expression (*i.e.*, expression that is at least two fold greater in a tumor than in normal tissue, as determined using a representative assay provided herein). Such screens may be performed, for example, using a Synteni microarray (Palo Alto,  
20 CA) according to the manufacturer's instructions (and essentially as described by Schena *et al.*, *Proc. Natl. Acad. Sci. USA* 93:10614-10619, 1996 and Heller *et al.*, *Proc. Natl. Acad. Sci. USA* 94:2150-2155, 1997). Alternatively, polynucleotides may be amplified from cDNA prepared from cells expressing the proteins described herein, such as breast tumor cells. Such polynucleotides may be amplified via polymerase  
25 chain reaction (PCR). For this approach, sequence-specific primers may be designed based on the sequences provided herein, and may be purchased or synthesized.

An amplified portion of a polynucleotide of the present invention may be used to isolate a full length gene from a suitable library (*e.g.*, a breast tumor cDNA library) using well known techniques. Within such techniques, a library (cDNA or  
30 genomic) is screened using one or more polynucleotide probes or primers suitable for

amplification. Preferably, a library is size-selected to include larger molecules. Random primed libraries may also be preferred for identifying 5' and upstream regions of genes. Genomic libraries are preferred for obtaining introns and extending 5' sequences.

5 For hybridization techniques, a partial sequence may be labeled (*e.g.*, by nick-translation or end-labeling with  $^{32}\text{P}$ ) using well known techniques. A bacterial or bacteriophage library is then generally screened by hybridizing filters containing denatured bacterial colonies (or lawns containing phage plaques) with the labeled probe (*see* Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor  
10 Laboratories, Cold Spring Harbor, NY, 1989). Hybridizing colonies or plaques are selected and expanded, and the DNA is isolated for further analysis. cDNA clones may be analyzed to determine the amount of additional sequence by, for example, PCR using a primer from the partial sequence and a primer from the vector. Restriction maps and partial sequences may be generated to identify one or more overlapping  
15 clones. The complete sequence may then be determined using standard techniques, which may involve generating a series of deletion clones. The resulting overlapping sequences can then be assembled into a single contiguous sequence. A full length cDNA molecule can be generated by ligating suitable fragments, using well known techniques.

Alternatively, there are numerous amplification techniques for obtaining  
20 a full length coding sequence from a partial cDNA sequence. Within such techniques, amplification is generally performed via PCR. Any of a variety of commercially available kits may be used to perform the amplification step. Primers may be designed using, for example, software well known in the art. Primers are preferably 22-30 nucleotides in length, have a GC content of at least 50% and anneal to the target  
25 sequence at temperatures of about 68°C to 72°C. The amplified region may be sequenced as described above, and overlapping sequences assembled into a contiguous sequence.

One such amplification technique is inverse PCR (*see* Triglia *et al.*, *Nucl. Acids Res.* 16:8186, 1988), which uses restriction enzymes to generate a fragment  
30 in the known region of the gene. The fragment is then circularized by intramolecular ligation and used as a template for PCR with divergent primers derived from the known

region. Within an alternative approach, sequences adjacent to a partial sequence may be retrieved by amplification with a primer to a linker sequence and a primer specific to a known region. The amplified sequences are typically subjected to a second round of amplification with the same linker primer and a second primer specific to the known  
5 region. A variation on this procedure, which employs two primers that initiate extension in opposite directions from the known sequence, is described in WO 96/38591. Another such technique is known as "rapid amplification of cDNA ends" or RACE. This technique involves the use of an internal primer and an external primer, which hybridizes to a polyA region or vector sequence, to identify sequences that are 5'  
10 and 3' of a known sequence. Additional techniques include capture PCR (Lagerstrom *et al.*, *PCR Methods Applic.* 1:111-19, 1991) and walking PCR (Parker *et al.*, *Nucl. Acids. Res.* 19:3055-60, 1991). Other methods employing amplification may also be employed to obtain a full length cDNA sequence.

In certain instances, it is possible to obtain a full length cDNA sequence  
15 by analysis of sequences provided in an expressed sequence tag (EST) database, such as that available from GenBank. Searches for overlapping ESTs may generally be performed using well known programs (*e.g.*, NCBI BLAST searches), and such ESTs may be used to generate a contiguous full length sequence. Full length DNA sequences may also be obtained by analysis of genomic fragments.

## 20 POLYNUCLEOTIDE EXPRESSION IN HOST CELLS

In other embodiments of the invention, polynucleotide sequences or fragments thereof which encode polypeptides of the invention, or fusion proteins or functional equivalents thereof, may be used in recombinant DNA molecules to direct expression of a polypeptide in appropriate host cells. Due to the inherent degeneracy of  
25 the genetic code, other DNA sequences that encode substantially the same or a functionally equivalent amino acid sequence may be produced and these sequences may be used to clone and express a given polypeptide.

As will be understood by those of skill in the art, it may be advantageous in some instances to produce polypeptide-encoding nucleotide sequences possessing  
30 non-naturally occurring codons. For example, codons preferred by a particular

prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

5                   Moreover, the polynucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter polypeptide encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. For example, DNA shuffling by random fragmentation and PCR reassembly of gene  
10 fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. In addition, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, or introduce mutations, and so forth.

                  In another embodiment of the invention, natural, modified, or  
15 recombinant nucleic acid sequences may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of polypeptide activity, it may be useful to encode a chimeric protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the polypeptide-encoding  
20 sequence and the heterologous protein sequence, so that the polypeptide may be cleaved and purified away from the heterologous moiety.

                  Sequences encoding a desired polypeptide may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M. H. *et al.* (1980) *Nucl. Acids Res. Symp. Ser.* 215-223, Horn, T. *et al.* (1980) *Nucl. Acids Res. Symp. Ser.* 225-232). Alternatively, the protein itself may be produced using chemical  
25 methods to synthesize the amino acid sequence of a polypeptide, or a portion thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J. Y. *et al.* (1995) *Science* 269:202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer, Palo  
30 Alto, CA).

A newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography (*e.g.*, Creighton, T. (1983) *Proteins, Structures and Molecular Principles*, WH Freeman and Co., New York, N.Y.) or other comparable techniques available in the art. The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (*e.g.*, the Edman degradation procedure). Additionally, the amino acid sequence of a polypeptide, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a desired polypeptide, the nucleotide sequences encoding the polypeptide, or functional equivalents, may be inserted into appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding a polypeptide of interest and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook, J. *et al.* (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview, N.Y., and Ausubel, F. M. *et al.* (1989) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, N.Y.

A variety of expression vector/host systems may be utilized to contain and express polynucleotide sequences. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (*e.g.*, baculovirus); plant cell systems transformed with virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (*e.g.*, Ti or pBR322 plasmids); or animal cell systems.

The "control elements" or "regulatory sequences" present in an expression vector are those non-translated regions of the vector--enhancers, promoters, 5' and 3' untranslated regions--which interact with host cellular proteins to carry out

transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid  
5 lacZ promoter of the PBLUESCRIPT phagemid (Stratagene, La Jolla, Calif.) or PSPORT1 plasmid (Gibco BRL, Gaithersburg, MD) and the like may be used. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are generally preferred. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding a polypeptide, vectors based on SV40 or EBV  
10 may be advantageously used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for the expressed polypeptide. For example, when large quantities are needed, for example for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used.  
15 Such vectors include, but are not limited to, the multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding the polypeptide of interest may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of .beta.-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S. M. Schuster (1989) *J.*  
20 *Biol. Chem.* 264:5503-5509); and the like. pGEX Vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to  
25 include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel *et al.* (supra) and Grant *et al.* (1987) *Methods*  
30 *Enzymol.* 153:516-544.



In cases where plant expression vectors are used, the expression of sequences encoding polypeptides may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. 5 (1987) *EMBO J.* 6:307-311. Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. *et al.* (1984) *EMBO J.* 3:1671-1680; Broglie, R. *et al.* (1984) *Science* 224:838-843; and Winter, J. *et al.* (1991) *Results Probl. Cell Differ.* 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques 10 are described in a number of generally available reviews (see, for example, Hobbs, S. or Murry, L. E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, N.Y.; pp. 191-196).

An insect system may also be used to express a polypeptide of interest. For example, in one such system, Autographa californica nuclear polyhedrosis virus 15 (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The sequences encoding the polypeptide may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of the polypeptide-encoding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat 20 protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia* larvae in which the polypeptide of interest may be expressed (Engelhard, E. K. *et al.* (1994) *Proc. Natl. Acad. Sci.* 91 :3224-3227).

In mammalian host cells, a number of viral-based expression systems are generally available. For example, in cases where an adenovirus is used as an expression 25 vector, sequences encoding a polypeptide of interest may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing the polypeptide in infected host cells (Logan, J. and Shenk, T. (1984) *Proc. Natl. Acad. Sci.* 81:3655-3659). In addition, 30 transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding a polypeptide of interest. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding the polypeptide, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. *et al.* (1994) *Results Probl. Cell Differ.* 20:125-162).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells such as CHO, HeLa, MDCK, HEK293, and WI38, which have specific cellular machinery and characteristic mechanisms for such post-translational activities, may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is generally preferred. For example, cell lines which stably express a polynucleotide of interest may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which

successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler, M. *et al.* (1977) *Cell* 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. *et al.* (1990) *Cell* 22:817-23) genes which can be employed in tk.sup.- or aprt.sup.- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler, M. *et al.* (1980) *Proc. Natl. Acad. Sci.* 77:3567-70); npt, which confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin, F. *et al.* (1981) *J. Mol. Biol.* 150:1-14); and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *supra*). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman, S. C. and R. C. Mulligan (1988) *Proc. Natl. Acad. Sci.* 85:8047-51). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C. A. *et al.* (1995) *Methods Mol. Biol.* 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding a polypeptide is inserted within a marker gene sequence, recombinant cells containing sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a polypeptide-encoding sequence under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain and express a desired polynucleotide sequence may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-

RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

A variety of protocols for detecting and measuring the expression of polynucleotide-encoded products, using either polyclonal or monoclonal antibodies specific for the product are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on a given polypeptide may be preferred for some applications, but a competitive binding assay may also be employed. These and other assays are described, among other places, in Hampton, R. *et al.* (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul. Minn.) and Maddox, D. E. *et al.* (1983; *J. Exp. Med.* 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences, or any portions thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits. Suitable reporter molecules or labels, which may be used include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with a polynucleotide sequence of interest may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides of the invention may be designed to contain signal sequences which direct secretion of the

encoded polypeptide through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may be used to join sequences encoding a polypeptide of interest to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, Wash.). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, Calif.) between the purification domain and the encoded polypeptide may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing a polypeptide of interest and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography) as described in Porath, J. *et al.* (1992, *Prot. Exp. Purif.* 3:263-281) while the enterokinase cleavage site provides a means for purifying the desired polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D. J. *et al.* (1993; *DNA Cell Biol.* 12:441-453).

In addition to recombinant production methods, polypeptides of the invention, and fragments thereof, may be produced by direct peptide synthesis using solid-phase techniques (Merrifield J. (1963) *J. Am. Chem. Soc.* 85:2149-2154). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Alternatively, various fragments may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

#### SITE-SPECIFIC MUTAGENESIS

Site-specific mutagenesis is a technique useful in the preparation of individual peptides, or biologically functional equivalent polypeptides, through specific mutagenesis of the underlying polynucleotides that encode them. The technique,

well-known to those of skill in the art, further provides a ready ability to prepare and test sequence variants, for example, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the DNA. Site-specific mutagenesis allows the production of mutants through the use of specific  
5 oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Mutations may be employed in a selected polynucleotide sequence to improve, alter, decrease, modify, or otherwise change the  
10 properties of the polynucleotide itself, and/or alter the properties, activity, composition, stability, or primary sequence of the encoded polypeptide.

In certain embodiments of the present invention, the inventors contemplate the mutagenesis of the disclosed polynucleotide sequences to alter one or more properties of the encoded polypeptide, such as the antigenicity of a polypeptide  
15 vaccine. The techniques of site-specific mutagenesis are well-known in the art, and are widely used to create variants of both polypeptides and polynucleotides. For example, site-specific mutagenesis is often used to alter a specific portion of a DNA molecule. In such embodiments, a primer comprising typically about 14 to about 25 nucleotides or so in length is employed, with about 5 to about 10 residues on both sides of the junction of  
20 the sequence being altered.

As will be appreciated by those of skill in the art, site-specific mutagenesis techniques have often employed a phage vector that exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage are readily  
25 commercially-available and their use is generally well-known to those skilled in the art. Double-stranded plasmids are also routinely employed in site directed mutagenesis that eliminates the step of transferring the gene of interest from a plasmid to a phage.

In general, site-directed mutagenesis in accordance herewith is performed by first obtaining a single-stranded vector or melting apart of two strands of  
30 a double-stranded vector that includes within its sequence a DNA sequence that encodes the desired peptide. An oligonucleotide primer bearing the desired mutated

sequence is prepared, generally synthetically. This primer is then annealed with the single-stranded vector, and subjected to DNA polymerizing enzymes such as *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original  
5 non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as *E. coli* cells, and clones are selected which include recombinant vectors bearing the mutated sequence arrangement.

The preparation of sequence variants of the selected peptide-encoding  
10 DNA segments using site-directed mutagenesis provides a means of producing potentially useful species and is not meant to be limiting as there are other ways in which sequence variants of peptides and the DNA sequences encoding them may be obtained. For example, recombinant vectors encoding the desired peptide sequence may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence  
15 variants. Specific details regarding these methods and protocols are found in the teachings of Maloy *et al.*, 1994; Segal, 1976; Prokop and Bajpai, 1991; Kuby, 1994; and Maniatis *et al.*, 1982, each incorporated herein by reference, for that purpose.

As used herein, the term "oligonucleotide directed mutagenesis procedure" refers to template-dependent processes and vector-mediated propagation  
20 which result in an increase in the concentration of a specific nucleic acid molecule relative to its initial concentration, or in an increase in the concentration of a detectable signal, such as amplification. As used herein, the term "oligonucleotide directed mutagenesis procedure" is intended to refer to a process that involves the template-dependent extension of a primer molecule. The term template dependent  
25 process refers to nucleic acid synthesis of an RNA or a DNA molecule wherein the sequence of the newly synthesized strand of nucleic acid is dictated by the well-known rules of complementary base pairing (see, for example, Watson, 1987). Typically, vector mediated methodologies involve the introduction of the nucleic acid fragment into a DNA or RNA vector, the clonal amplification of the vector, and the recovery of  
30 the amplified nucleic acid fragment. Examples of such methodologies are provided by U. S. Patent No. 4,237,224, specifically incorporated herein by reference in its entirety.

### POLYNUCLEOTIDE AMPLIFICATION TECHNIQUES

A number of template dependent processes are available to amplify the target sequences of interest present in a sample. One of the best known amplification methods is the polymerase chain reaction (PCR™) which is described in detail in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159, each of which is incorporated herein by reference in its entirety. Briefly, in PCR™, two primer sequences are prepared which are complementary to regions on opposite complementary strands of the target sequence. An excess of deoxynucleoside triphosphates is added to a reaction mixture along with a DNA polymerase (*e.g.*, *Taq* polymerase). If the target sequence is present in a sample, the primers will bind to the target and the polymerase will cause the primers to be extended along the target sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will dissociate from the target to form reaction products, excess primers will bind to the target and to the reaction product and the process is repeated. Preferably reverse transcription and PCR™ amplification procedure may be performed in order to quantify the amount of mRNA amplified. Polymerase chain reaction methodologies are well known in the art.

Another method for amplification is the ligase chain reaction (referred to as LCR), disclosed in Eur. Pat. Appl. Publ. No. 320,308 (specifically incorporated herein by reference in its entirety). In LCR, two complementary probe pairs are prepared, and in the presence of the target sequence, each pair will bind to opposite complementary strands of the target such that they abut. In the presence of a ligase, the two probe pairs will link to form a single unit. By temperature cycling, as in PCR™, bound ligated units dissociate from the target and then serve as "target sequences" for ligation of excess probe pairs. U.S. Patent No. 4,883,750, incorporated herein by reference in its entirety, describes an alternative method of amplification similar to LCR for binding probe pairs to a target sequence.

Qbeta Replicase, described in PCT Intl. Pat. Appl. Publ. No. PCT/US87/00880, incorporated herein by reference in its entirety, may also be used as still another amplification method in the present invention. In this method, a replicative sequence of RNA that has a region complementary to that of a target is added to a



sample in the presence of an RNA polymerase. The polymerase will copy the replicative sequence that can then be detected.

An isothermal amplification method, in which restriction endonucleases and ligases are used to achieve the amplification of target molecules that contain nucleotide 5'-[ $\alpha$ -thio]triphosphates in one strand of a restriction site (Walker *et al.*, 1992, incorporated herein by reference in its entirety), may also be useful in the amplification of nucleic acids in the present invention.

Strand Displacement Amplification (SDA) is another method of carrying out isothermal amplification of nucleic acids which involves multiple rounds of strand displacement and synthesis, *i.e.* nick translation. A similar method, called Repair Chain Reaction (RCR) is another method of amplification which may be useful in the present invention and is involves annealing several probes throughout a region targeted for amplification, followed by a repair reaction in which only two of the four bases are present. The other two bases can be added as biotinylated derivatives for easy detection. A similar approach is used in SDA.

Sequences can also be detected using a cyclic probe reaction (CPR). In CPR, a probe having a 3' and 5' sequences of non-target DNA and an internal or "middle" sequence of the target protein specific RNA is hybridized to DNA which is present in a sample. Upon hybridization, the reaction is treated with RNaseH, and the products of the probe are identified as distinctive products by generating a signal that is released after digestion. The original template is annealed to another cycling probe and the reaction is repeated. Thus, CPR involves amplifying a signal generated by hybridization of a probe to a target gene specific expressed nucleic acid.

Still other amplification methods described in Great Britain Pat. Appl. No. 2 202 328, and in PCT Intl. Pat. Appl. Publ. No. PCT/US89/01025, each of which is incorporated herein by reference in its entirety, may be used in accordance with the present invention. In the former application, "modified" primers are used in a PCR-like, template and enzyme dependent synthesis. The primers may be modified by labeling with a capture moiety (*e.g.*, biotin) and/or a detector moiety (*e.g.*, enzyme). In the latter application, an excess of labeled probes is added to a sample. In the presence of the target sequence, the probe binds and is cleaved catalytically. After cleavage, the

target sequence is released intact to be bound by excess probe. Cleavage of the labeled probe signals the presence of the target sequence.

Other nucleic acid amplification procedures include transcription-based amplification systems (TAS) (Kwoh *et al.*, 1989; PCT Intl. Pat. Appl. Publ. No. WO 88/10315, incorporated herein by reference in its entirety), including nucleic acid sequence based amplification (NASBA) and 3SR. In NASBA, the nucleic acids can be prepared for amplification by standard phenol/chloroform extraction, heat denaturation of a sample, treatment with lysis buffer and minispin columns for isolation of DNA and RNA or guanidinium chloride extraction of RNA. These amplification techniques involve annealing a primer that has sequences specific to the target sequence. Following polymerization, DNA/RNA hybrids are digested with RNase H while double stranded DNA molecules are heat-denatured again. In either case the single stranded DNA is made fully double stranded by addition of second target-specific primer, followed by polymerization. The double stranded DNA molecules are then multiply transcribed by a polymerase such as T7 or SP6. In an isothermal cyclic reaction, the RNAs are reverse transcribed into DNA, and transcribed once again with a polymerase such as T7 or SP6. The resulting products, whether truncated or complete, indicate target-specific sequences.

Eur. Pat. Appl. Publ. No. 329,822, incorporated herein by reference in its entirety, disclose a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA), which may be used in accordance with the present invention. The ssRNA is a first template for a first primer oligonucleotide, which is elongated by reverse transcriptase (RNA-dependent DNA polymerase). The RNA is then removed from resulting DNA:RNA duplex by the action of ribonuclease H (RNase H, an RNase specific for RNA in a duplex with either DNA or RNA). The resultant ssDNA is a second template for a second primer, which also includes the sequences of an RNA polymerase promoter (exemplified by T7 RNA polymerase) 5' to its homology to its template. This primer is then extended by DNA polymerase (exemplified by the large "Klenow" fragment of *E. coli* DNA polymerase I), resulting as a double-stranded DNA ("dsDNA") molecule, having a sequence identical to that of the original RNA between

the primers and having additionally, at one end, a promoter sequence. This promoter sequence can be used by the appropriate RNA polymerase to make many RNA copies of the DNA. These copies can then re-enter the cycle leading to very swift amplification. With proper choice of enzymes, this amplification can be done isothermally without addition of enzymes at each cycle. Because of the cyclical nature of this process, the starting sequence can be chosen to be in the form of either DNA or RNA.

PCT Intl. Pat. Appl. Publ. No. WO 89/06700, incorporated herein by reference in its entirety, disclose a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. This scheme is not cyclic; *i.e.* new templates are not produced from the resultant RNA transcripts. Other amplification methods include "RACE" (Frohman, 1990), and "one-sided PCR" (Ohara, 1989) which are well-known to those of skill in the art.

Methods based on ligation of two (or more) oligonucleotides in the presence of nucleic acid having the sequence of the resulting "di-oligonucleotide", thereby amplifying the di-oligonucleotide (Wu and Dean, 1996, incorporated herein by reference in its entirety), may also be used in the amplification of DNA sequences of the present invention.

## 20 BIOLOGICAL FUNCTIONAL EQUIVALENTS

Modification and changes may be made in the structure of the polynucleotides and polypeptides of the present invention and still obtain a functional molecule that encodes a polypeptide with desirable characteristics. As mentioned above, it is often desirable to introduce one or more mutations into a specific polynucleotide sequence. In certain circumstances, the resulting encoded polypeptide sequence is altered by this mutation, or in other cases, the sequence of the polypeptide is unchanged by one or more mutations in the encoding polynucleotide.

When it is desirable to alter the amino acid sequence of a polypeptide to create an equivalent, or even an improved, second-generation molecule, the amino acid

changes may be achieved by changing one or more of the codons of the encoding DNA sequence, according to Table 1.

For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence, and, of course, its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. It is thus contemplated by the inventors that various changes may be made in the peptide sequences of the disclosed compositions, or corresponding DNA sequences which encode said peptides without appreciable loss of their biological utility or activity.

**TABLE 1**

Amino Acids			Codons						
Alanine	Ala	A	GCA	GCC	GCG	GCU			
Cysteine	Cys	C	UGC	UGU					
Aspartic acid	Asp	D	GAC	GAU					
Glutamic acid	Glu	E	GAA	GAG					
Phenylalanine	Phe	F	UUC	UUU					
Glycine	Gly	G	GGA	GGC	GGG	GGU			
Histidine	His	H	CAC	CAU					
Isoleucine	Ile	I	AUA	AUC	AUU				
Lysine	Lys	K	AAA	AAG					
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU	
Methionine	Met	M	AUG						
Asparagine	Asn	N	AAC	AAU					
Proline	Pro	P	CCA	CCC	CCG	CCU			
Glutamine	Gln	Q	CAA	CAG					
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU	
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU	
Threonine	Thr	T	ACA	ACC	ACG	ACU			
Valine	Val	V	GUA	GUC	GUG	GUU			
Tryptophan	Trp	W	UGG						
Tyrosine	Tyr	Y	UAC	UAU					

In making such changes, the hydrophatic index of amino acids may be considered. The importance of the hydrophatic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982, incorporated herein by reference). It is accepted that the relative hydrophatic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like. Each amino acid has been assigned a hydrophatic index on the basis of its hydrophobicity and charge characteristics (Kyte and Doolittle, 1982). These values are:

isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydropathic index or score and still result in a protein with similar biological activity, *i.e.* still obtain a biological functionally equivalent protein. In making such changes, the substitution of amino acids whose hydropathic indices are within  $\pm 2$  is preferred, those within  $\pm 1$  are particularly preferred, and those within  $\pm 0.5$  are even more particularly preferred. It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U. S. Patent 4,554,101 (specifically incorporated herein by reference in its entirety), states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein.

As detailed in U. S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0  $\pm$  1); glutamate (+3.0  $\pm$  1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5  $\pm$  1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4). It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within  $\pm 2$  is preferred, those within  $\pm 1$  are particularly preferred, and those within  $\pm 0.5$  are even more particularly preferred.

As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions that take various of the foregoing characteristics into consideration are well known to those

of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

In addition, any polynucleotide may be further modified to increase stability *in vivo*. Possible modifications include, but are not limited to, the addition of  
5 flanking sequences at the 5' and/or 3' ends; the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages in the backbone; and/or the inclusion of nontraditional bases such as inosine, queosine and wybutosine, as well as acetyl-methyl-, thio- and other modified forms of adenine, cytidine, guanine, thymine and uridine.

## 10 IN VIVO POLYNUCLEOTIDE DELIVERY TECHNIQUES

In additional embodiments, genetic constructs comprising one or more of the polynucleotides of the invention are introduced into cells *in vivo*. This may be achieved using any of a variety of well known approaches, several of which are outlined below for the purpose of illustration.

### 15 1. ADENOVIRUS

One of the preferred methods for *in vivo* delivery of one or more nucleic acid sequences involves the use of an adenovirus expression vector. "Adenovirus expression vector" is meant to include those constructs containing adenovirus sequences sufficient to (a) support packaging of the construct and (b) to express a  
20 polynucleotide that has been cloned therein in a sense or antisense orientation. Of course, in the context of an antisense construct, expression does not require that the gene product be synthesized.

The expression vector comprises a genetically engineered form of an adenovirus. Knowledge of the genetic organization of adenovirus, a 36 kb, linear,  
25 double-stranded DNA virus, allows substitution of large pieces of adenoviral DNA with foreign sequences up to 7 kb (Grunhaus and Horwitz, 1992). In contrast to retrovirus, the adenoviral infection of host cells does not result in chromosomal integration because adenoviral DNA can replicate in an episomal manner without potential genotoxicity. Also, adenoviruses are structurally stable, and no genome rearrangement

has been detected after extensive amplification. Adenovirus can infect virtually all epithelial cells regardless of their cell cycle stage. So far, adenoviral infection appears to be linked only to mild disease such as acute respiratory disease in humans.

Adenovirus is particularly suitable for use as a gene transfer vector because of its mid-sized genome, ease of manipulation, high titer, wide target-cell range and high infectivity. Both ends of the viral genome contain 100-200 base pair inverted repeats (ITRs), which are *cis* elements necessary for viral DNA replication and packaging. The early (E) and late (L) regions of the genome contain different transcription units that are divided by the onset of viral DNA replication. The E1 region (E1A and E1B) encodes proteins responsible for the regulation of transcription of the viral genome and a few cellular genes. The expression of the E2 region (E2A and E2B) results in the synthesis of the proteins for viral DNA replication. These proteins are involved in DNA replication, late gene expression and host cell shut-off (Renan, 1990). The products of the late genes, including the majority of the viral capsid proteins, are expressed only after significant processing of a single primary transcript issued by the major late promoter (MLP). The MLP, (located at 16.8 m.u.) is particularly efficient during the late phase of infection, and all the mRNA's issued from this promoter possess a 5'-tripartite leader (TPL) sequence which makes them preferred mRNA's for translation.

In a current system, recombinant adenovirus is generated from homologous recombination between shuttle vector and provirus vector. Due to the possible recombination between two proviral vectors, wild-type adenovirus may be generated from this process. Therefore, it is critical to isolate a single clone of virus from an individual plaque and examine its genomic structure.

Generation and propagation of the current adenovirus vectors, which are replication deficient, depend on a unique helper cell line, designated 293, which was transformed from human embryonic kidney cells by Ad5 DNA fragments and constitutively expresses E1 proteins (Graham *et al.*, 1977). Since the E3 region is dispensable from the adenovirus genome (Jones and Shenk, 1978), the current adenovirus vectors, with the help of 293 cells, carry foreign DNA in either the E1, the D3 or both regions (Graham and Prevec, 1991). In nature, adenovirus can package



approximately 105% of the wild-type genome (Ghosh-Choudhury *et al.*, 1987), providing capacity for about 2 extra kB of DNA. Combined with the approximately 5.5 kB of DNA that is replaceable in the E1 and E3 regions, the maximum capacity of the current adenovirus vector is under 7.5 kB, or about 15% of the total length of the vector. More than 80% of the adenovirus viral genome remains in the vector backbone and is the source of vector-borne cytotoxicity. Also, the replication deficiency of the E1-deleted virus is incomplete. For example, leakage of viral gene expression has been observed with the currently available vectors at high multiplicities of infection (MOI) (Mulligan, 1993).

Helper cell lines may be derived from human cells such as human embryonic kidney cells, muscle cells, hematopoietic cells or other human embryonic mesenchymal or epithelial cells. Alternatively, the helper cells may be derived from the cells of other mammalian species that are permissive for human adenovirus. Such cells include, *e.g.*, Vero cells or other monkey embryonic mesenchymal or epithelial cells. As stated above, the currently preferred helper cell line is 293.

Recently, Racher *et al.* (1995) disclosed improved methods for culturing 293 cells and propagating adenovirus. In one format, natural cell aggregates are grown by inoculating individual cells into 1 liter siliconized spinner flasks (Techne, Cambridge, UK) containing 100-200 ml of medium. Following stirring at 40 rpm, the cell viability is estimated with trypan blue. In another format, Fibra-Cel microcarriers (Bibby Sterlin, Stone, UK) (5 g/l) is employed as follows. A cell inoculum, resuspended in 5 ml of medium, is added to the carrier (50 ml) in a 250 ml Erlenmeyer flask and left stationary, with occasional agitation, for 1 to 4 h. The medium is then replaced with 50 ml of fresh medium and shaking initiated. For virus production, cells are allowed to grow to about 80% confluence, after which time the medium is replaced (to 25% of the final volume) and adenovirus added at an MOI of 0.05. Cultures are left stationary overnight, following which the volume is increased to 100% and shaking commenced for another 72 h.

Other than the requirement that the adenovirus vector be replication defective, or at least conditionally defective, the nature of the adenovirus vector is not believed to be crucial to the successful practice of the invention. The adenovirus may

be of any of the 42 different known serotypes or subgroups A-F. Adenovirus type 5 of subgroup C is the preferred starting material in order to obtain a conditional replication-defective adenovirus vector for use in the present invention, since Adenovirus type 5 is a human adenovirus about which a great deal of biochemical and genetic information is known, and it has historically been used for most constructions employing adenovirus as a vector.

As stated above, the typical vector according to the present invention is replication defective and will not have an adenovirus E1 region. Thus, it will be most convenient to introduce the polynucleotide encoding the gene of interest at the position from which the E1-coding sequences have been removed. However, the position of insertion of the construct within the adenovirus sequences is not critical to the invention. The polynucleotide encoding the gene of interest may also be inserted in lieu of the deleted E3 region in E3 replacement vectors as described by Karlsson *et al.* (1986) or in the E4 region where a helper cell line or helper virus complements the E4 defect.

Adenovirus is easy to grow and manipulate and exhibits broad host range *in vitro* and *in vivo*. This group of viruses can be obtained in high titers, *e.g.*,  $10^9$ - $10^{11}$  plaque-forming units per ml, and they are highly infective. The life cycle of adenovirus does not require integration into the host cell genome. The foreign genes delivered by adenovirus vectors are episomal and, therefore, have low genotoxicity to host cells. No side effects have been reported in studies of vaccination with wild-type adenovirus (Couch *et al.*, 1963; Top *et al.*, 1971), demonstrating their safety and therapeutic potential as *in vivo* gene transfer vectors.

Adenovirus vectors have been used in eukaryotic gene expression (Levrero *et al.*, 1991; Gomez-Foix *et al.*, 1992) and vaccine development (Grunhaus and Horwitz, 1992; Graham and Prevec, 1992). Recently, animal studies suggested that recombinant adenovirus could be used for gene therapy (Stratford-Perricaudet and Perricaudet, 1991; Stratford-Perricaudet *et al.*, 1990; Rich *et al.*, 1993). Studies in administering recombinant adenovirus to different tissues include trachea instillation (Rosenfeld *et al.*, 1991; Rosenfeld *et al.*, 1992), muscle injection (Ragot *et al.*, 1993),

peripheral intravenous injections (Herz and Gerard, 1993) and stereotactic inoculation into the brain (Le Gal La Salle *et al.*, 1993).

## 2. RETROVIRUSES

The retroviruses are a group of single-stranded RNA viruses characterized by an ability to convert their RNA to double-stranded DNA in infected cells by a process of reverse-transcription (Coffin, 1990). The resulting DNA then stably integrates into cellular chromosomes as a provirus and directs synthesis of viral proteins. The integration results in the retention of the viral gene sequences in the recipient cell and its descendants. The retroviral genome contains three genes, gag, pol, and env that code for capsid proteins, polymerase enzyme, and envelope components, respectively. A sequence found upstream from the gag gene contains a signal for packaging of the genome into virions. Two long terminal repeat (LTR) sequences are present at the 5' and 3' ends of the viral genome. These contain strong promoter and enhancer sequences and are also required for integration in the host cell genome (Coffin, 1990).

In order to construct a retroviral vector, a nucleic acid encoding one or more oligonucleotide or polynucleotide sequences of interest is inserted into the viral genome in the place of certain viral sequences to produce a virus that is replication-defective. In order to produce virions, a packaging cell line containing the gag, pol, and env genes but without the LTR and packaging components is constructed (Mann *et al.*, 1983). When a recombinant plasmid containing a cDNA, together with the retroviral LTR and packaging sequences is introduced into this cell line (by calcium phosphate precipitation for example), the packaging sequence allows the RNA transcript of the recombinant plasmid to be packaged into viral particles, which are then secreted into the culture media (Nicolas and Rubenstein, 1988; Temin, 1986; Mann *et al.*, 1983). The media containing the recombinant retroviruses is then collected, optionally concentrated, and used for gene transfer. Retroviral vectors are able to infect a broad variety of cell types. However, integration and stable expression require the division of host cells (Paskind *et al.*, 1975).

A novel approach designed to allow specific targeting of retrovirus vectors was recently developed based on the chemical modification of a retrovirus by the chemical addition of lactose residues to the viral envelope. This modification could permit the specific infection of hepatocytes *via* sialoglycoprotein receptors.

5 A different approach to targeting of recombinant retroviruses was designed in which biotinylated antibodies against a retroviral envelope protein and against a specific cell receptor were used. The antibodies were coupled *via* the biotin components by using streptavidin (Roux *et al.*, 1989). Using antibodies against major histocompatibility complex class I and class II antigens, they demonstrated the infection  
10 of a variety of human cells that bore those surface antigens with an ecotropic virus *in vitro* (Roux *et al.*, 1989).

### 3. ADENO-ASSOCIATED VIRUSES

AAV (Ridgeway, 1988; Hermonat and Muzyczka, 1984) is a parovirus, discovered as a contamination of adenoviral stocks. It is a ubiquitous virus (antibodies  
15 are present in 85% of the US human population) that has not been linked to any disease. It is also classified as a dependovirus, because its replications is dependent on the presence of a helper virus, such as adenovirus. Five serotypes have been isolated, of which AAV-2 is the best characterized. AAV has a single-stranded linear DNA that is encapsidated into capsid proteins VP1, VP2 and VP3 to form an icosahedral virion of  
20 20 to 24 nm in diameter (Muzyczka and McLaughlin, 1988).

The AAV DNA is approximately 4.7 kilobases long. It contains two open reading frames and is flanked by two ITRs (FIG. 2). There are two major genes in the AAV genome: *rep* and *cap*. The *rep* gene codes for proteins responsible for viral replications, whereas *cap* codes for capsid protein VP1-3. Each ITR forms a T-shaped  
25 hairpin structure. These terminal repeats are the only essential *cis* components of the AAV for chromosomal integration. Therefore, the AAV can be used as a vector with all viral coding sequences removed and replaced by the cassette of genes for delivery. Three viral promoters have been identified and named p5, p19, and p40, according to their map position. Transcription from p5 and p19 results in production of rep proteins,

and transcription from p40 produces the capsid proteins (Hermonat and Muzyczka, 1984).

There are several factors that prompted researchers to study the possibility of using rAAV as an expression vector. One is that the requirements for  
5 delivering a gene to integrate into the host chromosome are surprisingly few. It is necessary to have the 145-bp ITRs, which are only 6% of the AAV genome. This leaves room in the vector to assemble a 4.5-kb DNA insertion. While this carrying capacity may prevent the AAV from delivering large genes, it is amply suited for delivering the antisense constructs of the present invention.

10 AAV is also a good choice of delivery vehicles due to its safety. There is a relatively complicated rescue mechanism: not only wild type adenovirus but also AAV genes are required to mobilize rAAV. Likewise, AAV is not pathogenic and not associated with any disease. The removal of viral coding sequences minimizes immune reactions to viral gene expression, and therefore, rAAV does not evoke an inflammatory  
15 response.

#### 4. OTHER VIRAL VECTORS AS EXPRESSION CONSTRUCTS

Other viral vectors may be employed as expression constructs in the present invention for the delivery of oligonucleotide or polynucleotide sequences to a host cell. Vectors derived from viruses such as vaccinia virus (Ridgeway, 1988;  
20 Coupar *et al.*, 1988), lentiviruses, polio viruses and herpes viruses may be employed. They offer several attractive features for various mammalian cells (Friedmann, 1989; Ridgeway, 1988; Coupar *et al.*, 1988; Horwich *et al.*, 1990).

With the recent recognition of defective hepatitis B viruses, new insight was gained into the structure-function relationship of different viral sequences. *In vitro*  
25 studies showed that the virus could retain the ability for helper-dependent packaging and reverse transcription despite the deletion of up to 80% of its genome (Horwich *et al.*, 1990). This suggested that large portions of the genome could be replaced with foreign genetic material. The hepatotropism and persistence (integration) were particularly attractive properties for liver-directed gene transfer. Chang *et al.* (1991)  
30 introduced the chloramphenicol acetyltransferase (CAT) gene into duck hepatitis B

virus genome in the place of the polymerase, surface, and pre-surface coding sequences. It was cotransfected with wild-type virus into an avian hepatoma cell line. Culture media containing high titers of the recombinant virus were used to infect primary duckling hepatocytes. Stable CAT gene expression was detected for at least 24 days after transfection (Chang *et al.*, 1991).

#### 5. NON-VIRAL VECTORS

In order to effect expression of the oligonucleotide or polynucleotide sequences of the present invention, the expression construct must be delivered into a cell. This delivery may be accomplished *in vitro*, as in laboratory procedures for transforming cells lines, or *in vivo* or *ex vivo*, as in the treatment of certain disease states. As described above, one preferred mechanism for delivery is *via* viral infection where the expression construct is encapsulated in an infectious viral particle.

Once the expression construct has been delivered into the cell the nucleic acid encoding the desired oligonucleotide or polynucleotide sequences may be positioned and expressed at different sites. In certain embodiments, the nucleic acid encoding the construct may be stably integrated into the genome of the cell. This integration may be in the specific location and orientation *via* homologous recombination (gene replacement) or it may be integrated in a random, non-specific location (gene augmentation). In yet further embodiments, the nucleic acid may be stably maintained in the cell as a separate, episomal segment of DNA. Such nucleic acid segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host cell cycle. How the expression construct is delivered to a cell and where in the cell the nucleic acid remains is dependent on the type of expression construct employed.

In certain embodiments of the invention, the expression construct comprising one or more oligonucleotide or polynucleotide sequences may simply consist of naked recombinant DNA or plasmids. Transfer of the construct may be performed by any of the methods mentioned above which physically or chemically permeabilize the cell membrane. This is particularly applicable for transfer *in vitro* but it may be applied to *in vivo* use as well. Dubensky *et al.* (1984) successfully injected

polyomavirus DNA in the form of calcium phosphate precipitates into liver and spleen of adult and newborn mice demonstrating active viral replication and acute infection. Benvenisty and Reshef (1986) also demonstrated that direct intraperitoneal injection of calcium phosphate-precipitated plasmids results in expression of the transfected genes.

- 5 It is envisioned that DNA encoding a gene of interest may also be transferred in a similar manner *in vivo* and express the gene product.

Another embodiment of the invention for transferring a naked DNA expression construct into cells may involve particle bombardment. This method depends on the ability to accelerate DNA-coated microprojectiles to a high velocity  
10 allowing them to pierce cell membranes and enter cells without killing them (Klein *et al.*, 1987). Several devices for accelerating small particles have been developed. One such device relies on a high voltage discharge to generate an electrical current, which in turn provides the motive force (Yang *et al.*, 1990). The microprojectiles used have consisted of biologically inert substances such as tungsten or gold beads.

- 15 Selected organs including the liver, skin, and muscle tissue of rats and mice have been bombarded *in vivo* (Yang *et al.*, 1990; Zelenin *et al.*, 1991). This may require surgical exposure of the tissue or cells, to eliminate any intervening tissue between the gun and the target organ, *i.e.* *ex vivo* treatment. Again, DNA encoding a particular gene may be delivered *via* this method and still be incorporated by the present  
20 invention.

#### ANTISENSE OLIGONUCLEOTIDES

- The end result of the flow of genetic information is the synthesis of protein. DNA is transcribed by polymerases into messenger RNA and translated on the ribosome to yield a folded, functional protein. Thus there are several steps along the  
25 route where protein synthesis can be inhibited. The native DNA segment coding for a polypeptide described herein, as all such mammalian DNA strands, has two strands: a sense strand and an antisense strand held together by hydrogen bonding. The messenger RNA coding for polypeptide has the same nucleotide sequence as the sense DNA strand except that the DNA thymidine is replaced by uridine. Thus, synthetic

antisense nucleotide sequences will bind to a mRNA and inhibit expression of the protein encoded by that mRNA.

The targeting of antisense oligonucleotides to mRNA is thus one mechanism to shut down protein synthesis, and, consequently, represents a powerful and targeted therapeutic approach. For example, the synthesis of polygalacturonase and the muscarine type 2 acetylcholine receptor are inhibited by antisense oligonucleotides directed to their respective mRNA sequences (U. S. Patent 5,739,119 and U. S. Patent 5,759,829, each specifically incorporated herein by reference in its entirety). Further, examples of antisense inhibition have been demonstrated with the nuclear protein cyclin, the multiple drug resistance gene (MDG1), ICAM-1, E-selectin, STK-1, striatal GABA<sub>A</sub> receptor and human EGF (Jaskulski *et al.*, 1988; Vasanthakumar and Ahmed, 1989; Peris *et al.*, 1998; U. S. Patent 5,801,154; U. S. Patent 5,789,573; U. S. Patent 5,718,709 and U. S. Patent 5,610,288, each specifically incorporated herein by reference in its entirety). Antisense constructs have also been described that inhibit and can be used to treat a variety of abnormal cellular proliferations, *e.g.* cancer (U. S. Patent 5,747,470; U. S. Patent 5,591,317 and U. S. Patent 5,783,683, each specifically incorporated herein by reference in its entirety).

Therefore, in exemplary embodiments, the invention provides oligonucleotide sequences that comprise all, or a portion of, any sequence that is capable of specifically binding to polynucleotide sequence described herein, or a complement thereof. In one embodiment, the antisense oligonucleotides comprise DNA or derivatives thereof. In another embodiment, the oligonucleotides comprise RNA or derivatives thereof. In a third embodiment, the oligonucleotides are modified DNAs comprising a phosphorothioated modified backbone. In a fourth embodiment, the oligonucleotide sequences comprise peptide nucleic acids or derivatives thereof. In each case, preferred compositions comprise a sequence region that is complementary, and more preferably substantially-complementary, and even more preferably, completely complementary to one or more portions of polynucleotides disclosed herein.

Selection of antisense compositions specific for a given gene sequence is based upon analysis of the chosen target sequence (*i.e.* in these illustrative examples the rat and human sequences) and determination of secondary structure,  $T_m$ , binding



energy, relative stability, and antisense compositions were selected based upon their relative inability to form dimers, hairpins, or other secondary structures that would reduce or prohibit specific binding to the target mRNA in a host cell.

Highly preferred target regions of the mRNA, are those which are at or  
5 near the AUG translation initiation codon, and those sequences which were substantially complementary to 5' regions of the mRNA. These secondary structure analyses and target site selection considerations were performed using v.4 of the OLIGO primer analysis software (Rychlik, 1997) and the BLASTN 2.0.5 algorithm software (Altschul *et al.*, 1997).

10 The use of an antisense delivery method employing a short peptide vector, termed MPG (27 residues), is also contemplated. The MPG peptide contains a hydrophobic domain derived from the fusion sequence of HIV gp41 and a hydrophilic domain from the nuclear localization sequence of SV40 T-antigen (Morris *et al.*, 1997). It has been demonstrated that several molecules of the MPG peptide coat the antisense  
15 oligonucleotides and can be delivered into cultured mammalian cells in less than 1 hour with relatively high efficiency (90%). Further, the interaction with MPG strongly increases both the stability of the oligonucleotide to nuclease and the ability to cross the plasma membrane (Morris *et al.*, 1997).

#### RIBOZYMES

20 Although proteins traditionally have been used for catalysis of nucleic acids, another class of macromolecules has emerged as useful in this endeavor. Ribozymes are RNA-protein complexes that cleave nucleic acids in a site-specific fashion. Ribozymes have specific catalytic domains that possess endonuclease activity (Kim and Cech, 1987; Gerlach *et al.*, 1987; Forster and Symons, 1987). For example, a  
25 large number of ribozymes accelerate phosphoester transfer reactions with a high degree of specificity, often cleaving only one of several phosphoesters in an oligonucleotide substrate (Cech *et al.*, 1981; Michel and Westhof, 1990; Reinhold-Hurek and Shub, 1992). This specificity has been attributed to the requirement that the substrate bind via specific base-pairing interactions to the internal guide sequence  
30 ("IGS") of the ribozyme prior to chemical reaction.

Ribozyme catalysis has primarily been observed as part of sequence-specific cleavage/ligation reactions involving nucleic acids (Joyce, 1989; Cech *et al.*, 1981). For example, U. S. Patent No. 5,354,855 (specifically incorporated herein by reference) reports that certain ribozymes can act as endonucleases with a sequence  
5 specificity greater than that of known ribonucleases and approaching that of the DNA restriction enzymes. Thus, sequence-specific ribozyme-mediated inhibition of gene expression may be particularly suited to therapeutic applications (Scanlon *et al.*, 1991; Sarver *et al.*, 1990). Recently, it was reported that ribozymes elicited genetic changes in some cells lines to which they were applied; the altered genes included the oncogenes  
10 H-ras, c-fos and genes of HIV. Most of this work involved the modification of a target mRNA, based on a specific mutant codon that is cleaved by a specific ribozyme.

Six basic varieties of naturally-occurring enzymatic RNAs are known presently. Each can catalyze the hydrolysis of RNA phosphodiester bonds *in trans* (and thus can cleave other RNA molecules) under physiological conditions. In general,  
15 enzymatic nucleic acids act by first binding to a target RNA. Such binding occurs through the target binding portion of a enzymatic nucleic acid which is held in close proximity to an enzymatic portion of the molecule that acts to cleave the target RNA. Thus, the enzymatic nucleic acid first recognizes and then binds a target RNA through complementary base-pairing, and once bound to the correct site, acts enzymatically to  
20 cut the target RNA. Strategic cleavage of such a target RNA will destroy its ability to direct synthesis of an encoded protein. After an enzymatic nucleic acid has bound and cleaved its RNA target, it is released from that RNA to search for another target and can repeatedly bind and cleave new targets.

The enzymatic nature of a ribozyme is advantageous over many  
25 technologies, such as antisense technology (where a nucleic acid molecule simply binds to a nucleic acid target to block its translation) since the concentration of ribozyme necessary to affect a therapeutic treatment is lower than that of an antisense oligonucleotide. This advantage reflects the ability of the ribozyme to act enzymatically. Thus, a single ribozyme molecule is able to cleave many molecules of  
30 target RNA. In addition, the ribozyme is a highly specific inhibitor, with the specificity of inhibition depending not only on the base pairing mechanism of binding to the target

RNA, but also on the mechanism of target RNA cleavage. Single mismatches, or base-substitutions, near the site of cleavage can completely eliminate catalytic activity of a ribozyme. Similar mismatches in antisense molecules do not prevent their action (Woolf *et al.*, 1992). Thus, the specificity of action of a ribozyme is greater than that of  
5 an antisense oligonucleotide binding the same RNA site.

The enzymatic nucleic acid molecule may be formed in a hammerhead, hairpin, a hepatitis  $\delta$  virus, group I intron or RNaseP RNA (in association with an RNA guide sequence) or Neurospora VS RNA motif. Examples of hammerhead motifs are described by Rossi *et al.* (1992). Examples of hairpin motifs are described by Hampel  
10 *et al.* (Eur. Pat. Appl. Publ. No. EP 0360257), Hampel and Tritz (1989), Hampel *et al.* (1990) and U. S. Patent 5,631,359 (specifically incorporated herein by reference). An example of the hepatitis  $\delta$  virus motif is described by Perrotta and Been (1992); an example of the RNaseP motif is described by Guerrier-Takada *et al.* (1983); Neurospora VS RNA ribozyme motif is described by Collins (Saville and Collins,  
15 1990; Saville and Collins, 1991; Collins and Olive, 1993); and an example of the Group I intron is described in (U. S. Patent 4,987,071, specifically incorporated herein by reference). All that is important in an enzymatic nucleic acid molecule of this invention is that it has a specific substrate binding site which is complementary to one or more of the target gene RNA regions, and that it have nucleotide sequences within or  
20 surrounding that substrate binding site which impart an RNA cleaving activity to the molecule. Thus the ribozyme constructs need not be limited to specific motifs mentioned herein.

In certain embodiments, it may be important to produce enzymatic cleaving agents which exhibit a high degree of specificity for the RNA of a desired  
25 target, such as one of the sequences disclosed herein. The enzymatic nucleic acid molecule is preferably targeted to a highly conserved sequence region of a target mRNA. Such enzymatic nucleic acid molecules can be delivered exogenously to specific cells as required. Alternatively, the ribozymes can be expressed from DNA or RNA vectors that are delivered to specific cells.

30 Small enzymatic nucleic acid motifs (*e.g.*, of the hammerhead or the hairpin structure) may also be used for exogenous delivery. The simple structure of

these molecules increases the ability of the enzymatic nucleic acid to invade targeted regions of the mRNA structure. Alternatively, catalytic RNA molecules can be expressed within cells from eukaryotic promoters (*e.g.*, Scanlon *et al.*, 1991; Kashani-Sabet *et al.*, 1992; Dropulic *et al.*, 1992; Weerasinghe *et al.*, 1991; Ojwang *et al.*, 1992; 5 Chen *et al.*, 1992; Sarver *et al.*, 1990). Those skilled in the art realize that any ribozyme can be expressed in eukaryotic cells from the appropriate DNA vector. The activity of such ribozymes can be augmented by their release from the primary transcript by a second ribozyme (Int. Pat. Appl. Publ. No. WO 93/23569, and Int. Pat. Appl. Publ. No. WO 94/02595, both hereby incorporated by reference; Ohkawa *et al.*, 10 1992; Taira *et al.*, 1991; and Ventura *et al.*, 1993).

Ribozymes may be added directly, or can be complexed with cationic lipids, lipid complexes, packaged within liposomes, or otherwise delivered to target cells. The RNA or RNA complexes can be locally administered to relevant tissues *ex vivo*, or *in vivo* through injection, aerosol inhalation, infusion pump or stent, with or 15 without their incorporation in biopolymers.

Ribozymes may be designed as described in Int. Pat. Appl. Publ. No. WO 93/23569 and Int. Pat. Appl. Publ. No. WO 94/02595, each specifically incorporated herein by reference) and synthesized to be tested *in vitro* and *in vivo*, as described. Such ribozymes can also be optimized for delivery. While specific 20 examples are provided, those in the art will recognize that equivalent RNA targets in other species can be utilized when necessary.

Hammerhead or hairpin ribozymes may be individually analyzed by computer folding (Jaeger *et al.*, 1989) to assess whether the ribozyme sequences fold into the appropriate secondary structure. Those ribozymes with unfavorable 25 intramolecular interactions between the binding arms and the catalytic core are eliminated from consideration. Varying binding arm lengths can be chosen to optimize activity. Generally, at least 5 or so bases on each arm are able to bind to, or otherwise interact with, the target RNA.

Ribozymes of the hammerhead or hairpin motif may be designed to 30 anneal to various sites in the mRNA message, and can be chemically synthesized. The method of synthesis used follows the procedure for normal RNA synthesis as described

in Usman *et al.* (1987) and in Scaringe *et al.* (1990) and makes use of common nucleic acid protecting and coupling groups, such as dimethoxytrityl at the 5'-end, and phosphoramidites at the 3'-end. Average stepwise coupling yields are typically >98%. Hairpin ribozymes may be synthesized in two parts and annealed to reconstruct an active ribozyme (Chowrira and Burke, 1992). Ribozymes may be modified extensively to enhance stability by modification with nuclease resistant groups, for example, 2'-amino, 2'-C-allyl, 2'-fluoro, 2'-o-methyl, 2'-H (for a review see *e.g.*, Usman and Cedergren, 1992). Ribozymes may be purified by gel electrophoresis using general methods or by high pressure liquid chromatography and resuspended in water.

10 Ribozyme activity can be optimized by altering the length of the ribozyme binding arms, or chemically synthesizing ribozymes with modifications that prevent their degradation by serum ribonucleases (see *e.g.*, Int. Pat. Appl. Publ. No. WO 92/07065; Perrault *et al.*, 1990; Pieken *et al.*, 1991; Usman and Cedergren, 1992; Int. Pat. Appl. Publ. No. WO 93/15187; Int. Pat. Appl. Publ. No. WO 91/03162; Eur. Pat. Appl. Publ. No. 92110298.4; U. S. Patent 5,334,711; and Int. Pat. Appl. Publ. No. WO 94/13688, which describe various chemical modifications that can be made to the sugar moieties of enzymatic RNA molecules), modifications which enhance their efficacy in cells, and removal of stem II bases to shorten RNA synthesis times and reduce chemical requirements.

20 Sullivan *et al.* (Int. Pat. Appl. Publ. No. WO 94/02595) describes the general methods for delivery of enzymatic RNA molecules. Ribozymes may be administered to cells by a variety of methods known to those familiar to the art, including, but not restricted to, encapsulation in liposomes, by iontophoresis, or by incorporation into other vehicles, such as hydrogels, cyclodextrins, biodegradable nanocapsules, and bioadhesive microspheres. For some indications, ribozymes may be directly delivered *ex vivo* to cells or tissues with or without the aforementioned vehicles. Alternatively, the RNA/vehicle combination may be locally delivered by direct inhalation, by direct injection or by use of a catheter, infusion pump or stent. Other routes of delivery include, but are not limited to, intravascular, intramuscular, subcutaneous or joint injection, aerosol inhalation, oral (tablet or pill form), topical, systemic, ocular, intraperitoneal and/or intrathecal delivery. More detailed descriptions

of ribozyme delivery and administration are provided in Int. Pat. Appl. Publ. No. WO 94/02595 and Int. Pat. Appl. Publ. No. WO 93/23569, each specifically incorporated herein by reference.

Another means of accumulating high concentrations of a ribozyme(s) within cells is to incorporate the ribozyme-encoding sequences into a DNA expression vector. Transcription of the ribozyme sequences are driven from a promoter for eukaryotic RNA polymerase I (pol I), RNA polymerase II (pol II), or RNA polymerase III (pol III). Transcripts from pol II or pol III promoters will be expressed at high levels in all cells; the levels of a given pol II promoter in a given cell type will depend on the nature of the gene regulatory sequences (enhancers, silencers, *etc.*) present nearby. Prokaryotic RNA polymerase promoters may also be used, providing that the prokaryotic RNA polymerase enzyme is expressed in the appropriate cells (Elroy-Stein and Moss, 1990; Gao and Huang, 1993; Lieber *et al.*, 1993; Zhou *et al.*, 1990). Ribozymes expressed from such promoters can function in mammalian cells (*e.g.* Kashani-Saber *et al.*, 1992; Ojwang *et al.*, 1992; Chen *et al.*, 1992; Yu *et al.*, 1993; L'Huillier *et al.*, 1992; Lisiewicz *et al.*, 1993). Such transcription units can be incorporated into a variety of vectors for introduction into mammalian cells, including but not restricted to, plasmid DNA vectors, viral DNA vectors (such as adenovirus or adeno-associated vectors), or viral RNA vectors (such as retroviral, semliki forest virus, sindbis virus vectors).

Ribozymes may be used as diagnostic tools to examine genetic drift and mutations within diseased cells. They can also be used to assess levels of the target RNA molecule. The close relationship between ribozyme activity and the structure of the target RNA allows the detection of mutations in any region of the molecule which alters the base-pairing and three-dimensional structure of the target RNA. By using multiple ribozymes, one may map nucleotide changes which are important to RNA structure and function *in vitro*, as well as in cells and tissues. Cleavage of target RNAs with ribozymes may be used to inhibit gene expression and define the role (essentially) of specified gene products in the progression of disease. In this manner, other genetic targets may be defined as important mediators of the disease. These studies will lead to better treatment of the disease progression by affording the possibility of combinational

therapies (e.g., multiple ribozymes targeted to different genes, ribozymes coupled with known small molecule inhibitors, or intermittent treatment with combinations of ribozymes and/or other chemical or biological molecules). Other *in vitro* uses of ribozymes are well known in the art, and include detection of the presence of mRNA associated with an IL-5 related condition. Such RNA is detected by determining the presence of a cleavage product after treatment with a ribozyme using standard methodology.

### PEPTIDE NUCLEIC ACIDS

In certain embodiments, the inventors contemplate the use of peptide nucleic acids (PNAs) in the practice of the methods of the invention. PNA is a DNA mimic in which the nucleobases are attached to a pseudopeptide backbone (Good and Nielsen, 1997). PNA is able to be utilized in a number methods that traditionally have used RNA or DNA. Often PNA sequences perform better in techniques than the corresponding RNA or DNA sequences and have utilities that are not inherent to RNA or DNA. A review of PNA including methods of making, characteristics of, and methods of using, is provided by Corey (1997) and is incorporated herein by reference. As such, in certain embodiments, one may prepare PNA sequences that are complementary to one or more portions of the ACE mRNA sequence, and such PNA compositions may be used to regulate, alter, decrease, or reduce the translation of ACE-specific mRNA, and thereby alter the level of ACE activity in a host cell to which such PNA compositions have been administered.

PNAs have 2-aminoethyl-glycine linkages replacing the normal phosphodiester backbone of DNA (Nielsen *et al.*, 1991; Hanvey *et al.*, 1992; Hyrup and Nielsen, 1996; Neilsen, 1996). This chemistry has three important consequences: firstly, in contrast to DNA or phosphorothioate oligonucleotides, PNAs are neutral molecules; secondly, PNAs are achiral, which avoids the need to develop a stereoselective synthesis; and thirdly, PNA synthesis uses standard Boc (Ducholm *et al.*, 1994) or Fmoc (Thomson *et al.*, 1995) protocols for solid-phase peptide synthesis, although other methods, including a modified Merrifield method, have been used (Christensen *et al.*, 1995).







PNA monomers or ready-made oligomers are commercially available from PerSeptive Biosystems (Framingham, MA). PNA syntheses by either Boc or Fmoc protocols are straightforward using manual or automated protocols (Norton *et al.*, 1995). The manual protocol lends itself to the production of chemically modified PNAs or the simultaneous synthesis of families of closely related PNAs.

As with peptide synthesis, the success of a particular PNA synthesis will depend on the properties of the chosen sequence. For example, while in theory PNAs can incorporate any combination of nucleotide bases, the presence of adjacent purines can lead to deletions of one or more residues in the product. In expectation of this difficulty, it is suggested that, in producing PNAs with adjacent purines, one should repeat the coupling of residues likely to be added inefficiently. This should be followed by the purification of PNAs by reverse-phase high-pressure liquid chromatography (Norton *et al.*, 1995) providing yields and purity of product similar to those observed during the synthesis of peptides.

Modifications of PNAs for a given application may be accomplished by coupling amino acids during solid-phase synthesis or by attaching compounds that contain a carboxylic acid group to the exposed N-terminal amine. Alternatively, PNAs can be modified after synthesis by coupling to an introduced lysine or cysteine. The ease with which PNAs can be modified facilitates optimization for better solubility or for specific functional requirements. Once synthesized, the identity of PNAs and their derivatives can be confirmed by mass spectrometry. Several studies have made and utilized modifications of PNAs (Norton *et al.*, 1995; Haaima *et al.*, 1996; Stetsenko *et al.*, 1996; Petersen *et al.*, 1995; Ulmann *et al.*, 1996; Koch *et al.*, 1995; Orum *et al.*, 1995; Footer *et al.*, 1996; Griffith *et al.*, 1995; Kremsky *et al.*, 1996; Pardridge *et al.*, 1995; Boffa *et al.*, 1995; Landsdorp *et al.*, 1996; Gambacorti-Passerini *et al.*, 1996; Armitage *et al.*, 1997; Seeger *et al.*, 1997; Ruskowski *et al.*, 1997). U.S. Patent No. 5,700,922 discusses PNA-DNA-PNA chimeric molecules and their uses in diagnostics, modulating protein in organisms, and treatment of conditions susceptible to therapeutics.

In contrast to DNA and RNA, which contain negatively charged linkages, the PNA backbone is neutral. In spite of this dramatic alteration, PNAs

recognize complementary DNA and RNA by Watson-Crick pairing (Egholm *et al.*, 1993), validating the initial modeling by Nielsen *et al.* (1991). PNAs lack 3' to 5' polarity and can bind in either parallel or antiparallel fashion, with the antiparallel mode being preferred (Egholm *et al.*, 1993).

5 Hybridization of DNA oligonucleotides to DNA and RNA is destabilized by electrostatic repulsion between the negatively charged phosphate backbones of the complementary strands. By contrast, the absence of charge repulsion in PNA-DNA or PNA-RNA duplexes increases the melting temperature ( $T_m$ ) and reduces the dependence of  $T_m$  on the concentration of mono- or divalent cations  
10 (Nielsen *et al.*, 1991). The enhanced rate and affinity of hybridization are significant because they are responsible for the surprising ability of PNAs to perform strand invasion of complementary sequences within relaxed double-stranded DNA. In addition, the efficient hybridization at inverted repeats suggests that PNAs can recognize secondary structure effectively within double-stranded DNA. Enhanced  
15 recognition also occurs with PNAs immobilized on surfaces, and Wang *et al.* have shown that support-bound PNAs can be used to detect hybridization events (Wang *et al.*, 1996).

One might expect that tight binding of PNAs to complementary sequences would also increase binding to similar (but not identical) sequences, reducing  
20 the sequence specificity of PNA recognition. As with DNA hybridization, however, selective recognition can be achieved by balancing oligomer length and incubation temperature. Moreover, selective hybridization of PNAs is encouraged by PNA-DNA hybridization being less tolerant of base mismatches than DNA-DNA hybridization. For example, a single mismatch within a 16 bp PNA-DNA duplex can reduce the  $T_m$  by  
25 up to 15°C (Egholm *et al.*, 1993). This high level of discrimination has allowed the development of several PNA-based strategies for the analysis of point mutations (Wang *et al.*, 1996; Carlsson *et al.*, 1996; Thiede *et al.*, 1996; Webb and Hurskainen, 1996; Perry-O'Keefe *et al.*, 1996).

High-affinity binding provides clear advantages for molecular  
30 recognition and the development of new applications for PNAs. For example, 11-13 nucleotide PNAs inhibit the activity of telomerase, a ribonucleo-protein that extends

telomere ends using an essential RNA template, while the analogous DNA oligomers do not (Norton *et al.*, 1996).

Neutral PNAs are more hydrophobic than analogous DNA oligomers, and this can lead to difficulty solubilizing them at neutral pH, especially if the PNAs have a high purine content or if they have the potential to form secondary structures. Their solubility can be enhanced by attaching one or more positive charges to the PNA termini (Nielsen *et al.*, 1991).

Findings by Allfrey and colleagues suggest that strand invasion will occur spontaneously at sequences within chromosomal DNA (Boffa *et al.*, 1995; Boffa *et al.*, 1996). These studies targeted PNAs to triplet repeats of the nucleotides CAG and used this recognition to purify transcriptionally active DNA (Boffa *et al.*, 1995) and to inhibit transcription (Boffa *et al.*, 1996). This result suggests that if PNAs can be delivered within cells then they will have the potential to be general sequence-specific regulators of gene expression. Studies and reviews concerning the use of PNAs as antisense and anti-gene agents include Nielsen *et al.* (1993b), Hanvey *et al.* (1992), and Good and Nielsen (1997). Koppelhus *et al.* (1997) have used PNAs to inhibit HIV-1 inverse transcription, showing that PNAs may be used for antiviral therapies.

Methods of characterizing the antisense binding properties of PNAs are discussed in Rose (1993) and Jensen *et al.* (1997). Rose uses capillary gel electrophoresis to determine binding of PNAs to their complementary oligonucleotide, measuring the relative binding kinetics and stoichiometry. Similar types of measurements were made by Jensen *et al.* using BIAcore™ technology.

Other applications of PNAs include use in DNA strand invasion (Nielsen *et al.*, 1991), antisense inhibition (Hanvey *et al.*, 1992), mutational analysis (Orum *et al.*, 1993), enhancers of transcription (Mollegaard *et al.*, 1994), nucleic acid purification (Orum *et al.*, 1995), isolation of transcriptionally active genes (Boffa *et al.*, 1995), blocking of transcription factor binding (Vickers *et al.*, 1995), genome cleavage (Veselkov *et al.*, 1996), biosensors (Wang *et al.*, 1996), *in situ* hybridization (Thisted *et al.*, 1996), and in a alternative to Southern blotting (Perry-O'Keefe, 1996).

### POLYPEPTIDE COMPOSITIONS

The present invention, in other aspects, provides polypeptide compositions. Generally, a polypeptide of the invention will be an isolated polypeptide (or an epitope, variant, or active fragment thereof) derived from a mammalian species.

- 5 Preferably, the polypeptide is encoded by a polynucleotide sequence disclosed herein or a sequence which hybridizes under moderately stringent conditions to a polynucleotide sequence disclosed herein. Alternatively, the polypeptide may be defined as a polypeptide which comprises a contiguous amino acid sequence from an amino acid sequence disclosed herein, or which polypeptide comprises an entire amino acid  
10 sequence disclosed herein.

- In the present invention, a polypeptide composition is also understood to comprise one or more polypeptides that are immunologically reactive with antibodies generated against a polypeptide of the invention, particularly a polypeptide having the amino acid sequence disclosed in SEQ ID NO:126, or to active fragments, or to variants  
15 or biological functional equivalents thereof.

- Likewise, a polypeptide composition of the present invention is understood to comprise one or more polypeptides that are capable of eliciting antibodies that are immunologically reactive with one or more polypeptides encoded by one or more contiguous nucleic acid sequences contained in SEQ ID NO:1-125, or to active  
20 fragments, or to variants thereof, or to one or more nucleic acid sequences which hybridize to one or more of these sequences under conditions of moderate to high stringency. Particularly illustrative polypeptides include the amino acid sequence disclosed in SEQ ID NO:126.

- As used herein, an active fragment of a polypeptide includes a whole or  
25 a portion of a polypeptide which is modified by conventional techniques, *e.g.*, mutagenesis, or by addition, deletion, or substitution, but which active fragment exhibits substantially the same structure function, antigenicity, etc., as a polypeptide as described herein.

- In certain illustrative embodiments, the polypeptides of the invention  
30 will comprise at least an immunogenic portion of a breast tumor protein or a variant thereof, as described herein. As noted above, a "breast tumor protein" is a protein that

is expressed by breast tumor cells. Proteins that are breast tumor proteins also react detectably within an immunoassay (such as an ELISA) with antisera from a patient with breast cancer. Polypeptides as described herein may be of any length. Additional sequences derived from the native protein and/or heterologous sequences may be present, and such sequences may (but need not) possess further immunogenic or antigenic properties.

An "immunogenic portion," as used herein is a portion of a protein that is recognized (*i.e.*, specifically bound) by a B-cell and/or T-cell surface antigen receptor. Such immunogenic portions generally comprise at least 5 amino acid residues, more preferably at least 10, and still more preferably at least 20 amino acid residues of a breast tumor protein or a variant thereof. Certain preferred immunogenic portions include peptides in which an N-terminal leader sequence and/or transmembrane domain have been deleted. Other preferred immunogenic portions may contain a small N- and/or C-terminal deletion (*e.g.*, 1-30 amino acids, preferably 5-15 amino acids), relative to the mature protein.

Immunogenic portions may generally be identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3rd ed., 243-247 (Raven Press, 1993) and references cited therein. Such techniques include screening polypeptides for the ability to react with antigen-specific antibodies, antisera and/or T-cell lines or clones. As used herein, antisera and antibodies are "antigen-specific" if they specifically bind to an antigen (*i.e.*, they react with the protein in an ELISA or other immunoassay, and do not react detectably with unrelated proteins). Such antisera and antibodies may be prepared as described herein, and using well known techniques. An immunogenic portion of a native breast tumor protein is a portion that reacts with such antisera and/or T-cells at a level that is not substantially less than the reactivity of the full length polypeptide (*e.g.*, in an ELISA and/or T-cell reactivity assay). Such immunogenic portions may react within such assays at a level that is similar to or greater than the reactivity of the full length polypeptide. Such screens may generally be performed using methods well known to those of ordinary skill in the art, such as those described in Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. For example, a polypeptide may be

immobilized on a solid support and contacted with patient sera to allow binding of antibodies within the sera to the immobilized polypeptide. Unbound sera may then be removed and bound antibodies detected using, for example, <sup>125</sup>I-labeled Protein A.

As noted above, a composition may comprise a variant of a native breast tumor protein. A polypeptide "variant," as used herein, is a polypeptide that differs from a native breast tumor protein in one or more substitutions, deletions, additions and/or insertions, such that the immunogenicity of the polypeptide is not substantially diminished. In other words, the ability of a variant to react with antigen-specific antisera may be enhanced or unchanged, relative to the native protein, or may be diminished by less than 50%, and preferably less than 20%, relative to the native protein. Such variants may generally be identified by modifying one of the above polypeptide sequences and evaluating the reactivity of the modified polypeptide with antigen-specific antibodies or antisera as described herein. Preferred variants include those in which one or more portions, such as an N-terminal leader sequence or transmembrane domain, have been removed. Other preferred variants include variants in which a small portion (*e.g.*, 1-30 amino acids, preferably 5-15 amino acids) has been removed from the N- and/or C-terminal of the mature protein.

Polypeptide variants encompassed by the present invention include those exhibiting at least about 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more identity (determined as described above) to the polypeptides disclosed herein.

Preferably, a variant contains conservative substitutions. A "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydrophobic nature of the polypeptide to be substantially unchanged. Amino acid substitutions may generally be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity and/or the amphipathic nature of the residues. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine and valine; glycine and

alanine; asparagine and glutamine; and serine, threonine, phenylalanine and tyrosine. Other groups of amino acids that may represent conservative changes include: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his. A variant may also, or alternatively, contain  
5 nonconservative changes. In a preferred embodiment, variant polypeptides differ from a native sequence by substitution, deletion or addition of five amino acids or fewer. Variants may also (or alternatively) be modified by, for example, the deletion or addition of amino acids that have minimal influence on the immunogenicity, secondary structure and hydrophobic nature of the polypeptide.

10 As noted above, polypeptides may comprise a signal (or leader) sequence at the N-terminal end of the protein, which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (e.g., poly-His), or to enhance binding of the polypeptide to a solid support.  
15 For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

Polypeptides may be prepared using any of a variety of well known techniques. Recombinant polypeptides encoded by DNA sequences as described above may be readily prepared from the DNA sequences using any of a variety of expression vectors known to those of ordinary skill in the art. Expression may be achieved in any  
20 appropriate host cell that has been transformed or transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable host cells include prokaryotes, yeast, and higher eukaryotic cells, such as mammalian cells and plant cells. Preferably, the host cells employed are *E. coli*, yeast or a mammalian cell line such as COS or CHO. Supernatants from suitable host/vector systems which  
25 secrete recombinant protein or polypeptide into culture media may be first concentrated using a commercially available filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an affinity matrix or an ion exchange resin. Finally, one or more reverse phase HPLC steps can be employed to further purify a recombinant polypeptide.

30 Portions and other variants having less than about 100 amino acids, and generally less than about 50 amino acids, may also be generated by synthetic means,



using techniques well known to those of ordinary skill in the art. For example, such polypeptides may be synthesized using any of the commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. See Merrifield, *J. Am. Chem. Soc.* 85:2149-2146, 1963. Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Perkin Elmer/Applied BioSystems Division (Foster City, CA), and may be operated according to the manufacturer's instructions.

Within certain specific embodiments, a polypeptide may be a fusion protein that comprises multiple polypeptides as described herein, or that comprises at least one polypeptide as described herein and an unrelated sequence, such as a known tumor protein. A fusion partner may, for example, assist in providing T helper epitopes (an immunological fusion partner), preferably T helper epitopes recognized by humans, or may assist in expressing the protein (an expression enhancer) at higher yields than the native recombinant protein. Certain preferred fusion partners are both immunological and expression enhancing fusion partners. Other fusion partners may be selected so as to increase the solubility of the protein or to enable the protein to be targeted to desired intracellular compartments. Still further fusion partners include affinity tags, which facilitate purification of the protein.

Fusion proteins may generally be prepared using standard techniques, including chemical conjugation. Preferably, a fusion protein is expressed as a recombinant protein, allowing the production of increased levels, relative to a non-fused protein, in an expression system. Briefly, DNA sequences encoding the polypeptide components may be assembled separately, and ligated into an appropriate expression vector. The 3' end of the DNA sequence encoding one polypeptide component is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide component so that the reading frames of the sequences are in phase. This permits translation into a single fusion protein that retains the biological activity of both component polypeptides.

A peptide linker sequence may be employed to separate the first and second polypeptide components by a distance sufficient to ensure that each polypeptide

5 folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion protein using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as  
10 linkers include those disclosed in Maratea *et al.*, *Gene* 40:39-46, 1985; Murphy *et al.*, *Proc. Natl. Acad. Sci. USA* 83:8258-8262, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may generally be from 1 to about 50 amino acids in length. Linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to  
15 separate the functional domains and prevent steric interference.

The ligated DNA sequences are operably linked to suitable transcriptional or translational regulatory elements. The regulatory elements responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptides. Similarly, stop codons required to end translation and  
20 transcription termination signals are only present 3' to the DNA sequence encoding the second polypeptide.

Fusion proteins are also provided. Such proteins comprise a polypeptide as described herein together with an unrelated immunogenic protein. Preferably the immunogenic protein is capable of eliciting a recall response. Examples of such  
25 proteins include tetanus, tuberculosis and hepatitis proteins (*see*, for example, Stoute *et al.* *New Engl. J. Med.*, 336:86-91, 1997).

Within preferred embodiments, an immunological fusion partner is derived from protein D, a surface protein of the gram-negative bacterium *Haemophilus influenza B* (WO 91/18926). Preferably, a protein D derivative comprises  
30 approximately the first third of the protein (*e.g.*, the first N-terminal 100-110 amino acids), and a protein D derivative may be lipidated. Within certain preferred

embodiments, the first 109 residues of a Lipoprotein D fusion partner is included on the N-terminus to provide the polypeptide with additional exogenous T-cell epitopes and to increase the expression level in *E. coli* (thus functioning as an expression enhancer). The lipid tail ensures optimal presentation of the antigen to antigen presenting cells.

5 Other fusion partners include the non-structural protein from influenzae virus, NS1 (hemagglutinin). Typically, the N-terminal 81 amino acids are used, although different fragments that include T-helper epitopes may be used.

In another embodiment, the immunological fusion partner is the protein known as LYTA, or a portion thereof (preferably a C-terminal portion). LYTA is derived from *Streptococcus pneumoniae*, which synthesizes an N-acetyl-L-alanine

10 amidase known as amidase LYTA (encoded by the *LytA* gene; *Gene* 43:265-292, 1986). LYTA is an autolysin that specifically degrades certain bonds in the peptidoglycan backbone. The C-terminal domain of the LYTA protein is responsible for the affinity to the choline or to some choline analogues such as DEAE. This

15 property has been exploited for the development of *E. coli* C-LYTA expressing plasmids useful for expression of fusion proteins. Purification of hybrid proteins containing the C-LYTA fragment at the amino terminus has been described (see *Biotechnology* 10:795-798, 1992). Within a preferred embodiment, a repeat portion of LYTA may be incorporated into a fusion protein. A repeat portion is found in the C-

20 terminal region starting at residue 178. A particularly preferred repeat portion incorporates residues 188-305.

In general, polypeptides (including fusion proteins) and polynucleotides as described herein are isolated. An "isolated" polypeptide or polynucleotide is one that is removed from its original environment. For example, a naturally-occurring protein is

25 isolated if it is separated from some or all of the coexisting materials in the natural system. Preferably, such polypeptides are at least about 90% pure, more preferably at least about 95% pure and most preferably at least about 99% pure. A polynucleotide is considered to be isolated if, for example, it is cloned into a vector that is not a part of the natural environment.

**BINDING AGENTS**

The present invention further provides agents, such as antibodies and antigen-binding fragments thereof, that specifically bind to a breast tumor protein. As used herein, an antibody, or antigen-binding fragment thereof, is said to "specifically bind" to a breast tumor protein if it reacts at a detectable level (within, for example, an ELISA) with a breast tumor protein, and does not react detectably with unrelated proteins under similar conditions. As used herein, "binding" refers to a noncovalent association between two separate molecules such that a complex is formed. The ability to bind may be evaluated by, for example, determining a binding constant for the formation of the complex. The binding constant is the value obtained when the concentration of the complex is divided by the product of the component concentrations. In general, two compounds are said to "bind," in the context of the present invention, when the binding constant for complex formation exceeds about  $10^3$  L/mol. The binding constant may be determined using methods well known in the art.

Binding agents may be further capable of differentiating between patients with and without a cancer, such as breast cancer, using the representative assays provided herein. In other words, antibodies or other binding agents that bind to a breast tumor protein will generate a signal indicating the presence of a cancer in at least about 20% of patients with the disease, and will generate a negative signal indicating the absence of the disease in at least about 90% of individuals without the cancer. To determine whether a binding agent satisfies this requirement, biological samples (*e.g.*, blood, sera, sputum, urine and/or tumor biopsies) from patients with and without a cancer (as determined using standard clinical tests) may be assayed as described herein for the presence of polypeptides that bind to the binding agent. It will be apparent that a statistically significant number of samples with and without the disease should be assayed. Each binding agent should satisfy the above criteria; however, those of ordinary skill in the art will recognize that binding agents may be used in combination to improve sensitivity.

Any agent that satisfies the above requirements may be a binding agent. For example, a binding agent may be a ribosome, with or without a peptide component, an RNA molecule or a polypeptide. In a preferred embodiment, a binding agent is an

antibody or an antigen-binding fragment thereof. Antibodies may be prepared by any of a variety of techniques known to those of ordinary skill in the art. See, e.g., Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In general, antibodies can be produced by cell culture techniques, including the generation  
5 of monoclonal antibodies as described herein, or via transfection of antibody genes into suitable bacterial or mammalian cell hosts, in order to allow for the production of recombinant antibodies. In one technique, an immunogen comprising the polypeptide is initially injected into any of a wide variety of mammals (e.g., mice, rats, rabbits, sheep or goats). In this step, the polypeptides of this invention may serve as the  
10 immunogen without modification. Alternatively, particularly for relatively short polypeptides, a superior immune response may be elicited if the polypeptide is joined to a carrier protein, such as bovine serum albumin or keyhole limpet hemocyanin. The immunogen is injected into the animal host, preferably according to a predetermined schedule incorporating one or more booster immunizations, and the animals are bled  
15 periodically. Polyclonal antibodies specific for the polypeptide may then be purified from such antisera by, for example, affinity chromatography using the polypeptide coupled to a suitable solid support.

Monoclonal antibodies specific for an antigenic polypeptide of interest may be prepared, for example, using the technique of Kohler and Milstein, *Eur. J. Immunol.* 6:511-519, 1976, and improvements thereto. Briefly, these methods involve  
20 the preparation of immortal cell lines capable of producing antibodies having the desired specificity (i.e., reactivity with the polypeptide of interest). Such cell lines may be produced, for example, from spleen cells obtained from an animal immunized as described above. The spleen cells are then immortalized by, for example, fusion with a  
25 myeloma cell fusion partner, preferably one that is syngeneic with the immunized animal. A variety of fusion techniques may be employed. For example, the spleen cells and myeloma cells may be combined with a nonionic detergent for a few minutes and then plated at low density on a selective medium that supports the growth of hybrid cells, but not myeloma cells. A preferred selection technique uses HAT (hypoxanthine,  
30 aminopterin, thymidine) selection. After a sufficient time, usually about 1 to 2 weeks, colonies of hybrids are observed. Single colonies are selected and their culture

supernatants tested for binding activity against the polypeptide. Hybridomas having high reactivity and specificity are preferred.

Monoclonal antibodies may be isolated from the supernatants of growing hybridoma colonies. In addition, various techniques may be employed to enhance the yield, such as injection of the hybridoma cell line into the peritoneal cavity of a suitable vertebrate host, such as a mouse. Monoclonal antibodies may then be harvested from the ascites fluid or the blood. Contaminants may be removed from the antibodies by conventional techniques, such as chromatography, gel filtration, precipitation, and extraction. The polypeptides of this invention may be used in the purification process in, for example, an affinity chromatography step.

Within certain embodiments, the use of antigen-binding fragments of antibodies may be preferred. Such fragments include Fab fragments, which may be prepared using standard techniques. Briefly, immunoglobulins may be purified from rabbit serum by affinity chromatography on Protein A bead columns (Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988) and digested by papain to yield Fab and Fc fragments. The Fab and Fc fragments may be separated by affinity chromatography on protein A bead columns.

Monoclonal antibodies of the present invention may be coupled to one or more therapeutic agents. Suitable agents in this regard include radionuclides, differentiation inducers, drugs, toxins, and derivatives thereof. Preferred radionuclides include  $^{90}\text{Y}$ ,  $^{123}\text{I}$ ,  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{186}\text{Re}$ ,  $^{188}\text{Re}$ ,  $^{211}\text{At}$ , and  $^{212}\text{Bi}$ . Preferred drugs include methotrexate, and pyrimidine and purine analogs. Preferred differentiation inducers include phorbol esters and butyric acid. Preferred toxins include ricin, abrin, diphtheria toxin, cholera toxin, gelonin, *Pseudomonas* exotoxin, *Shigella* toxin, and pokeweed antiviral protein.

A therapeutic agent may be coupled (e.g., covalently bonded) to a suitable monoclonal antibody either directly or indirectly (e.g., via a linker group). A direct reaction between an agent and an antibody is possible when each possesses a substituent capable of reacting with the other. For example, a nucleophilic group, such as an amino or sulfhydryl group, on one may be capable of reacting with a carbonyl-

containing group, such as an anhydride or an acid halide, or with an alkyl group containing a good leaving group (e.g., a halide) on the other.

Alternatively, it may be desirable to couple a therapeutic agent and an antibody via a linker group. A linker group can function as a spacer to distance an  
5 antibody from an agent in order to avoid interference with binding capabilities. A linker group can also serve to increase the chemical reactivity of a substituent on an agent or an antibody, and thus increase the coupling efficiency. An increase in chemical reactivity may also facilitate the use of agents, or functional groups on agents, which otherwise would not be possible.

10 It will be evident to those skilled in the art that a variety of bifunctional or polyfunctional reagents, both homo- and hetero-functional (such as those described in the catalog of the Pierce Chemical Co., Rockford, IL), may be employed as the linker group. Coupling may be effected, for example, through amino groups, carboxyl groups, sulfhydryl groups or oxidized carbohydrate residues. There are numerous references  
15 describing such methodology, e.g., U.S. Patent No. 4,671,958, to Rodwell *et al.*

Where a therapeutic agent is more potent when free from the antibody portion of the immunoconjugates of the present invention, it may be desirable to use a linker group which is cleavable during or upon internalization into a cell. A number of different cleavable linker groups have been described. The mechanisms for the  
20 intracellular release of an agent from these linker groups include cleavage by reduction of a disulfide bond (e.g., U.S. Patent No. 4,489,710, to Spitler), by irradiation of a photolabile bond (e.g., U.S. Patent No. 4,625,014, to Senter *et al.*), by hydrolysis of derivatized amino acid side chains (e.g., U.S. Patent No. 4,638,045, to Kohn *et al.*), by serum complement-mediated hydrolysis (e.g., U.S. Patent No. 4,671,958, to Rodwell *et al.*), and acid-catalyzed hydrolysis (e.g., U.S. Patent No. 4,569,789, to Blattler *et al.*).  
25

It may be desirable to couple more than one agent to an antibody. In one embodiment, multiple molecules of an agent are coupled to one antibody molecule. In another embodiment, more than one type of agent may be coupled to one antibody. Regardless of the particular embodiment, immunoconjugates with more than one agent  
30 may be prepared in a variety of ways. For example, more than one agent may be

coupled directly to an antibody molecule, or linkers that provide multiple sites for attachment can be used. Alternatively, a carrier can be used.

A carrier may bear the agents in a variety of ways, including covalent bonding either directly or via a linker group. Suitable carriers include proteins such as albumins (*e.g.*, U.S. Patent No. 4,507,234, to Kato *et al.*), peptides and polysaccharides such as aminodextran (*e.g.*, U.S. Patent No. 4,699,784, to Shih *et al.*). A carrier may also bear an agent by noncovalent bonding or by encapsulation, such as within a liposome vesicle (*e.g.*, U.S. Patent Nos. 4,429,008 and 4,873,088). Carriers specific for radionuclide agents include radiohalogenated small molecules and chelating compounds. For example, U.S. Patent No. 4,735,792 discloses representative radiohalogenated small molecules and their synthesis. A radionuclide chelate may be formed from chelating compounds that include those containing nitrogen and sulfur atoms as the donor atoms for binding the metal, or metal oxide, radionuclide. For example, U.S. Patent No. 4,673,562, to Davison *et al.* discloses representative chelating compounds and their synthesis.

A variety of routes of administration for the antibodies and immunoconjugates may be used. Typically, administration will be intravenous, intramuscular, subcutaneous or in the bed of a resected tumor. It will be evident that the precise dose of the antibody/immunoconjugate will vary depending upon the antibody used, the antigen density on the tumor, and the rate of clearance of the antibody.

#### T CELLS

Immunotherapeutic compositions may also, or alternatively, comprise T cells specific for a breast tumor protein. Such cells may generally be prepared *in vitro* or *ex vivo*, using standard procedures. For example, T cells may be isolated from bone marrow, peripheral blood, or a fraction of bone marrow or peripheral blood of a patient, using a commercially available cell separation system, such as the Isolex™ System, available from Nexell Therapeutics, Inc. (Irvine, CA; see also U.S. Patent No. 5,240,856; U.S. Patent No. 5,215,926; WO 89/06280; WO 91/16116 and WO



92/07243). Alternatively, T cells may be derived from related or unrelated humans, non-human mammals, cell lines or cultures.

T cells may be stimulated with a breast tumor polypeptide, polynucleotide encoding a breast tumor polypeptide and/or an antigen presenting cell (APC) that expresses such a polypeptide. Such stimulation is performed under conditions and for a time sufficient to permit the generation of T cells that are specific for the polypeptide. Preferably, a breast tumor polypeptide or polynucleotide is present within a delivery vehicle, such as a microsphere, to facilitate the generation of specific T cells.

T cells are considered to be specific for a breast tumor polypeptide if the T cells specifically proliferate, secrete cytokines or kill target cells coated with the polypeptide or expressing a gene encoding the polypeptide. T cell specificity may be evaluated using any of a variety of standard techniques. For example, within a chromium release assay or proliferation assay, a stimulation index of more than two fold increase in lysis and/or proliferation, compared to negative controls, indicates T cell specificity. Such assays may be performed, for example, as described in Chen *et al.*, *Cancer Res.* 54:1065-1070, 1994. Alternatively, detection of the proliferation of T cells may be accomplished by a variety of known techniques. For example, T cell proliferation can be detected by measuring an increased rate of DNA synthesis (*e.g.*, by pulse-labeling cultures of T cells with tritiated thymidine and measuring the amount of tritiated thymidine incorporated into DNA). Contact with a breast tumor polypeptide (100 ng/ml - 100 µg/ml, preferably 200 ng/ml - 25 µg/ml) for 3 - 7 days should result in at least a two fold increase in proliferation of the T cells. Contact as described above for 2-3 hours should result in activation of the T cells, as measured using standard cytokine assays in which a two fold increase in the level of cytokine release (*e.g.*, TNF or IFN-γ) is indicative of T cell activation (*see* Coligan *et al.*, *Current Protocols in Immunology*, vol. 1, Wiley Interscience (Greene 1998)). T cells that have been activated in response to a breast tumor polypeptide, polynucleotide or polypeptide-expressing APC may be CD4<sup>+</sup> and/or CD8<sup>+</sup>. Breast tumor protein-specific T cells may be expanded using standard techniques. Within preferred embodiments, the T cells are

derived from a patient, a related donor or an unrelated donor, and are administered to the patient following stimulation and expansion.

For therapeutic purposes, CD4<sup>+</sup> or CD8<sup>+</sup> T cells that proliferate in response to a breast tumor polypeptide, polynucleotide or APC can be expanded in number either *in vitro* or *in vivo*. Proliferation of such T cells *in vitro* may be accomplished in a variety of ways. For example, the T cells can be re-exposed to a breast tumor polypeptide, or a short peptide corresponding to an immunogenic portion of such a polypeptide, with or without the addition of T cell growth factors, such as interleukin-2, and/or stimulator cells that synthesize a breast tumor polypeptide. Alternatively, one or more T cells that proliferate in the presence of a breast tumor protein can be expanded in number by cloning. Methods for cloning cells are well known in the art, and include limiting dilution.

#### PHARMACEUTICAL COMPOSITIONS

In additional embodiments, the present invention concerns formulation of one or more of the polynucleotide, polypeptide, T-cell and/or antibody compositions disclosed herein in pharmaceutically-acceptable solutions for administration to a cell or an animal, either alone, or in combination with one or more other modalities of therapy.

It will also be understood that, if desired, the nucleic acid segment, RNA, DNA or PNA compositions that express a polypeptide as disclosed herein may be administered in combination with other agents as well, such as, *e.g.*, other proteins or polypeptides or various pharmaceutically-active agents. In fact, there is virtually no limit to other components that may also be included, given that the additional agents do not cause a significant adverse effect upon contact with the target cells or host tissues. The compositions may thus be delivered along with various other agents as required in the particular instance. Such compositions may be purified from host cells or other biological sources, or alternatively may be chemically synthesized as described herein. Likewise, such compositions may further comprise substituted or derivatized RNA or DNA compositions.

Formulation of pharmaceutically-acceptable excipients and carrier solutions is well-known to those of skill in the art, as is the development of suitable

dosing and treatment regimens for using the particular compositions described herein in a variety of treatment regimens, including *e.g.*, oral, parenteral, intravenous, intranasal, and intramuscular administration and formulation.

#### 1. ORAL DELIVERY

- 5                   In certain applications, the pharmaceutical compositions disclosed herein may be delivered *via* oral administration to an animal. As such, these compositions may be formulated with an inert diluent or with an assimilable edible carrier, or they may be enclosed in hard- or soft-shell gelatin capsule, or they may be compressed into tablets, or they may be incorporated directly with the food of the diet.
- 10                   The active compounds may even be incorporated with excipients and used in the form of ingestible tablets, buccal tables, troches, capsules, elixirs, suspensions, syrups, wafers, and the like (Mathiowitz *et al.*, 1997; Hwang *et al.*, 1998; U. S. Patent 5,641,515; U. S. Patent 5,580,579 and U. S. Patent 5,792,451, each specifically incorporated herein by reference in its entirety). The tablets, troches, pills,
- 15 capsules and the like may also contain the following: a binder, as gum tragacanth, acacia, cornstarch, or gelatin; excipients, such as dicalcium phosphate; a disintegrating agent, such as corn starch, potato starch, alginic acid and the like; a lubricant, such as magnesium stearate; and a sweetening agent, such as sucrose, lactose or saccharin may be added or a flavoring agent, such as peppermint, oil of wintergreen, or cherry
- 20 flavoring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar, or both. A syrup or elixir may contain the active compound sucrose as a sweetening agent methyl and
- 25 propylparabens as preservatives, a dye and flavoring, such as cherry or orange flavor. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compounds may be incorporated into sustained-release preparation and formulations.

Typically, these formulations may contain at least about 0.1% of the active compound or more, although the percentage of the active ingredient(s) may, of course, be varied and may conveniently be between about 1 or 2% and about 60% or 70% or more of the weight or volume of the total formulation. Naturally, the amount of active compound(s) in each therapeutically useful composition may be prepared in such a way that a suitable dosage will be obtained in any given unit dose of the compound. Factors such as solubility, bioavailability, biological half-life, route of administration, product shelf life, as well as other pharmacological considerations will be contemplated by one skilled in the art of preparing such pharmaceutical formulations, and as such, a variety of dosages and treatment regimens may be desirable.

For oral administration the compositions of the present invention may alternatively be incorporated with one or more excipients in the form of a mouthwash, dentifrice, buccal tablet, oral spray, or sublingual orally-administered formulation. For example, a mouthwash may be prepared incorporating the active ingredient in the required amount in an appropriate solvent, such as a sodium borate solution (Dobell's Solution). Alternatively, the active ingredient may be incorporated into an oral solution such as one containing sodium borate, glycerin and potassium bicarbonate, or dispersed in a dentifrice, or added in a therapeutically-effective amount to a composition that may include water, binders, abrasives, flavoring agents, foaming agents, and humectants. Alternatively the compositions may be fashioned into a tablet or solution form that may be placed under the tongue or otherwise dissolved in the mouth.

## 2. INJECTABLE DELIVERY

In certain circumstances it will be desirable to deliver the pharmaceutical compositions disclosed herein parenterally, intravenously, intramuscularly, or even intraperitoneally as described in U. S. Patent 5,543,158; U. S. Patent 5,641,515 and U. S. Patent 5,399,363 (each specifically incorporated herein by reference in its entirety). Solutions of the active compounds as free base or pharmacologically acceptable salts may be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions may also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of

storage and use, these preparations contain a preservative to prevent the growth of microorganisms.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions (U. S. Patent 5,466,468, specifically incorporated herein by reference in its entirety). In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (*e.g.*, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and/or vegetable oils. Proper fluidity may be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The prevention of the action of microorganisms can be facilitated by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous and intraperitoneal administration. In this connection, a sterile aqueous medium that can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage may be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event,

determine the appropriate dose for the individual subject. Moreover, for human administration, preparations should meet sterility, pyrogenicity, and the general safety and purity standards as required by FDA Office of Biologics standards.

5 Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the  
10 preparation of sterile injectable solutions, the preferred methods of preparation are vacuum-drying and freeze-drying techniques which yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

The compositions disclosed herein may be formulated in a neutral or salt  
15 form. Pharmaceutically-acceptable salts, include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium,  
20 ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like. Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms such as injectable solutions, drug-release capsules, and the like.

25 As used herein, "carrier" includes any and all solvents, dispersion media, vehicles, coatings, diluents, antibacterial and antifungal agents, isotonic and absorption delaying agents, buffers, carrier solutions, suspensions, colloids, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active  
30 ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

The phrase "pharmaceutically-acceptable" refers to molecular entities and compositions that do not produce an allergic or similar untoward reaction when administered to a human. The preparation of an aqueous composition that contains a protein as an active ingredient is well understood in the art. Typically, such  
5 compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection can also be prepared. The preparation can also be emulsified.

### 3. NASAL DELIVERY

In certain embodiments, the pharmaceutical compositions may be  
10 delivered by intranasal sprays, inhalation, and/or other aerosol delivery vehicles. Methods for delivering genes, nucleic acids, and peptide compositions directly to the lungs *via* nasal aerosol sprays has been described *e.g.*, in U. S. Patent 5,756,353 and U. S. Patent 5,804,212 (each specifically incorporated herein by reference in its entirety). Likewise, the delivery of drugs using intranasal microparticle resins (Takenaga *et al.*,  
15 1998) and lysophosphatidyl-glycerol compounds (U. S. Patent 5,725,871, specifically incorporated herein by reference in its entirety) are also well-known in the pharmaceutical arts. Likewise, transmucosal drug delivery in the form of a polytetrafluoroethylene support matrix is described in U. S. Patent 5,780,045 (specifically incorporated herein by reference in its entirety).

### 20 4. LIPOSOME-, NANOCAPSULE-, AND MICROPARTICLE-MEDIATED DELIVERY

In certain embodiments, the inventors contemplate the use of liposomes, nanocapsules, microparticles, microspheres, lipid particles, vesicles, and the like, for the introduction of the compositions of the present invention into suitable host cells. In particular, the compositions of the present invention may be formulated for delivery  
25 either encapsulated in a lipid particle, a liposome, a vesicle, a nanosphere, or a nanoparticle or the like.

Such formulations may be preferred for the introduction of pharmaceutically-acceptable formulations of the nucleic acids or constructs disclosed herein. The formation and use of liposomes is generally known to those of skill in the

art (see for example, Couvreur *et al.*, 1977; Couvreur, 1988; Lasic, 1998; which describes the use of liposomes and nanocapsules in the targeted antibiotic therapy for intracellular bacterial infections and diseases). Recently, liposomes were developed with improved serum stability and circulation half-times (Gabizon and  
5 Papahadjopoulos, 1988; Allen and Choun, 1987; U. S. Patent 5,741,516, specifically incorporated herein by reference in its entirety). Further, various methods of liposome and liposome like preparations as potential drug carriers have been reviewed (Takakura, 1998; Chandran *et al.*, 1997; Margalit, 1995; U. S. Patent 5,567,434; U. S. Patent 5,552,157; U. S. Patent 5,565,213; U. S. Patent 5,738,868 and U. S. Patent 5,795,587,  
10 each specifically incorporated herein by reference in its entirety).

Liposomes have been used successfully with a number of cell types that are normally resistant to transfection by other procedures including T cell suspensions, primary hepatocyte cultures and PC 12 cells (Renneisen *et al.*, 1990; Muller *et al.*, 1990). In addition, liposomes are free of the DNA length constraints that are typical of  
15 viral-based delivery systems. Liposomes have been used effectively to introduce genes, drugs (Heath and Martin, 1986; Heath *et al.*, 1986; Balazsovits *et al.*, 1989; Fresta and Puglisi, 1996), radiotherapeutic agents (Pikul *et al.*, 1987), enzymes (Imaizumi *et al.*, 1990a; Imaizumi *et al.*, 1990b), viruses (Faller and Baltimore, 1984), transcription factors and allosteric effectors (Nicolau and Gersonde, 1979) into a variety of cultured  
20 cell lines and animals. In addition, several successful clinical trials examining the effectiveness of liposome-mediated drug delivery have been completed (Lopez-Berestein *et al.*, 1985a; 1985b; Coune, 1988; Sculier *et al.*, 1988). Furthermore, several studies suggest that the use of liposomes is not associated with autoimmune responses, toxicity or gonadal localization after systemic delivery (Mori and Fukatsu, 1992).

25 Liposomes are formed from phospholipids that are dispersed in an aqueous medium and spontaneously form multilamellar concentric bilayer vesicles (also termed multilamellar vesicles (MLVs). MLVs generally have diameters of from 25 nm to 4  $\mu$ m. Sonication of MLVs results in the formation of small unilamellar vesicles (SUVs) with diameters in the range of 200 to 500 Å, containing an aqueous  
30 solution in the core.



Liposomes bear resemblance to cellular membranes and are contemplated for use in connection with the present invention as carriers for the peptide compositions. They are widely suitable as both water- and lipid-soluble substances can be entrapped, *i.e.* in the aqueous spaces and within the bilayer itself, respectively. It is possible that the drug-bearing liposomes may even be employed for site-specific delivery of active agents by selectively modifying the liposomal formulation.

In addition to the teachings of Couvreur *et al.* (1977; 1988), the following information may be utilized in generating liposomal formulations. Phospholipids can form a variety of structures other than liposomes when dispersed in water, depending on the molar ratio of lipid to water. At low ratios the liposome is the preferred structure. The physical characteristics of liposomes depend on pH, ionic strength and the presence of divalent cations. Liposomes can show low permeability to ionic and polar substances, but at elevated temperatures undergo a phase transition which markedly alters their permeability. The phase transition involves a change from a closely packed, ordered structure, known as the gel state, to a loosely packed, less-ordered structure, known as the fluid state. This occurs at a characteristic phase-transition temperature and results in an increase in permeability to ions, sugars and drugs.

In addition to temperature, exposure to proteins can alter the permeability of liposomes. Certain soluble proteins, such as cytochrome c, bind, deform and penetrate the bilayer, thereby causing changes in permeability. Cholesterol inhibits this penetration of proteins, apparently by packing the phospholipids more tightly. It is contemplated that the most useful liposome formations for antibiotic and inhibitor delivery will contain cholesterol.

The ability to trap solutes varies between different types of liposomes. For example, MLVs are moderately efficient at trapping solutes, but SUVs are extremely inefficient. SUVs offer the advantage of homogeneity and reproducibility in size distribution, however, and a compromise between size and trapping efficiency is offered by large unilamellar vesicles (LUVs). These are prepared by ether evaporation and are three to four times more efficient at solute entrapment than MLVs.

In addition to liposome characteristics, an important determinant in entrapping compounds is the physicochemical properties of the compound itself. Polar compounds are trapped in the aqueous spaces and nonpolar compounds bind to the lipid bilayer of the vesicle. Polar compounds are released through permeation or when the bilayer is broken, but nonpolar compounds remain affiliated with the bilayer unless it is disrupted by temperature or exposure to lipoproteins. Both types show maximum efflux rates at the phase transition temperature.

Liposomes interact with cells *via* four different mechanisms: endocytosis by phagocytic cells of the reticuloendothelial system such as macrophages and neutrophils; adsorption to the cell surface, either by nonspecific weak hydrophobic or electrostatic forces, or by specific interactions with cell-surface components; fusion with the plasma cell membrane by insertion of the lipid bilayer of the liposome into the plasma membrane, with simultaneous release of liposomal contents into the cytoplasm; and by transfer of liposomal lipids to cellular or subcellular membranes, or vice versa, without any association of the liposome contents. It often is difficult to determine which mechanism is operative and more than one may operate at the same time.

The fate and disposition of intravenously injected liposomes depend on their physical properties, such as size, fluidity, and surface charge. They may persist in tissues for h or days, depending on their composition, and half lives in the blood range from min to several h. Larger liposomes, such as MLVs and LUVs, are taken up rapidly by phagocytic cells of the reticuloendothelial system, but physiology of the circulatory system restrains the exit of such large species at most sites. They can exit only in places where large openings or pores exist in the capillary endothelium, such as the sinusoids of the liver or spleen. Thus, these organs are the predominate site of uptake. On the other hand, SUVs show a broader tissue distribution but still are sequestered highly in the liver and spleen. In general, this *in vivo* behavior limits the potential targeting of liposomes to only those organs and tissues accessible to their large size. These include the blood, liver, spleen, bone marrow, and lymphoid organs.

Targeting is generally not a limitation in terms of the present invention. However, should specific targeting be desired, methods are available for this to be accomplished. Antibodies may be used to bind to the liposome surface and to direct the

antibody and its drug contents to specific antigenic receptors located on a particular cell-type surface. Carbohydrate determinants (glycoprotein or glycolipid cell-surface components that play a role in cell-cell recognition, interaction and adhesion) may also be used as recognition sites as they have potential in directing liposomes to particular  
5 cell types. Mostly, it is contemplated that intravenous injection of liposomal preparations would be used, but other routes of administration are also conceivable.

Alternatively, the invention provides for pharmaceutically-acceptable nanocapsule formulations of the compositions of the present invention. Nanocapsules can generally entrap compounds in a stable and reproducible way (Henry-Michelland  
10 *et al.*, 1987; Quintanar-Guerrero *et al.*, 1998; Douglas *et al.*, 1987). To avoid side effects due to intracellular polymeric overloading, such ultrafine particles (sized around 0.1  $\mu\text{m}$ ) should be designed using polymers able to be degraded *in vivo*. Biodegradable polyalkyl-cyanoacrylate nanoparticles that meet these requirements are contemplated for use in the present invention. Such particles may be easily made, as described  
15 (Couvreux *et al.*, 1980; 1988; zur Muhlen *et al.*, 1998; Zambaux *et al.* 1998; Pinto-Alphandry *et al.*, 1995 and U. S. Patent 5,145,684, specifically incorporated herein by reference in its entirety).

#### VACCINES

In certain preferred embodiments of the present invention, vaccines are  
20 provided. The vaccines will generally comprise one or more pharmaceutical compositions, such as those discussed above, in combination with an immunostimulant. An immunostimulant may be any substance that enhances or potentiates an immune response (antibody and/or cell-mediated) to an exogenous antigen. Examples of immunostimulants include adjuvants, biodegradable microspheres (*e.g.*, polylactic  
25 galactide) and liposomes (into which the compound is incorporated; *see e.g.*, Fullerton, U.S. Patent No. 4,235,877). Vaccine preparation is generally described in, for example, M.F. Powell and M.J. Newman, eds., "Vaccine Design (the subunit and adjuvant approach)," Plenum Press (NY, 1995). Pharmaceutical compositions and vaccines within the scope of the present invention may also contain other compounds, which  
30 may be biologically active or inactive. For example, one or more immunogenic

portions of other tumor antigens may be present, either incorporated into a fusion polypeptide or as a separate compound, within the composition or vaccine.

Illustrative vaccines may contain DNA encoding one or more of the polypeptides as described above, such that the polypeptide is generated *in situ*. As noted above, the DNA may be present within any of a variety of delivery systems known to those of ordinary skill in the art, including nucleic acid expression systems, bacteria and viral expression systems. Numerous gene delivery techniques are well known in the art, such as those described by Rolland, *Crit. Rev. Therap. Drug Carrier Systems* 15:143-198, 1998, and references cited therein. Appropriate nucleic acid expression systems contain the necessary DNA sequences for expression in the patient (such as a suitable promoter and terminating signal). Bacterial delivery systems involve the administration of a bacterium (such as *Bacillus-Calmette-Guerrin*) that expresses an immunogenic portion of the polypeptide on its cell surface or secretes such an epitope. In a preferred embodiment, the DNA may be introduced using a viral expression system (e.g., vaccinia or other pox virus, retrovirus, or adenovirus), which may involve the use of a non-pathogenic (defective), replication competent virus. Suitable systems are disclosed, for example, in Fisher-Hoch *et al.*, *Proc. Natl. Acad. Sci. USA* 86:317-321, 1989; Flexner *et al.*, *Ann. N.Y. Acad. Sci.* 569:86-103, 1989; Flexner *et al.*, *Vaccine* 8:17-21, 1990; U.S. Patent Nos. 4,603,112, 4,769,330, and 5,017,487; WO 89/01973; U.S. Patent No. 4,777,127; GB 2,200,651; EP 0,345,242; WO 91/02805; Berkner, *Biotechniques* 6:616-627, 1988; Rosenfeld *et al.*, *Science* 252:431-434, 1991; Kolls *et al.*, *Proc. Natl. Acad. Sci. USA* 91:215-219, 1994; Kass-Eisler *et al.*, *Proc. Natl. Acad. Sci. USA* 90:11498-11502, 1993; Guzman *et al.*, *Circulation* 88:2838-2848, 1993; and Guzman *et al.*, *Cir. Res.* 73:1202-1207, 1993. Techniques for incorporating DNA into such expression systems are well known to those of ordinary skill in the art. The DNA may also be "naked," as described, for example, in Ulmer *et al.*, *Science* 259:1745-1749, 1993 and reviewed by Cohen, *Science* 259:1691-1692, 1993. The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells. It will be apparent that a vaccine may comprise both a polynucleotide and a polypeptide component. Such vaccines may provide for an enhanced immune response.

It will be apparent that a vaccine may contain pharmaceutically acceptable salts of the polynucleotides and polypeptides provided herein. Such salts may be prepared from pharmaceutically acceptable non-toxic bases, including organic bases (*e.g.*, salts of primary, secondary and tertiary amines and basic amino acids) and  
5 inorganic bases (*e.g.*, sodium, potassium, lithium, ammonium, calcium and magnesium salts).

While any suitable carrier known to those of ordinary skill in the art may be employed in the vaccine compositions of this invention, the type of carrier will vary depending on the mode of administration. Compositions of the present invention may  
10 be formulated for any appropriate manner of administration, including for example, topical, oral, nasal, intravenous, intracranial, intraperitoneal, subcutaneous or intramuscular administration. For parenteral administration, such as subcutaneous injection, the carrier preferably comprises water, saline, alcohol, a fat, a wax or a buffer. For oral administration, any of the above carriers or a solid carrier, such as mannitol,  
15 lactose, starch, magnesium stearate, sodium saccharine, talcum, cellulose, glucose, sucrose, and magnesium carbonate, may be employed. Biodegradable microspheres (*e.g.*, polylactate polyglycolate) may also be employed as carriers for the pharmaceutical compositions of this invention. Suitable biodegradable microspheres are disclosed, for example, in U.S. Patent Nos. 4,897,268; 5,075,109; 5,928,647;  
20 5,811,128; 5,820,883; 5,853,763; 5,814,344 and 5,942,252. One may also employ a carrier comprising the particulate-protein complexes described in U.S. Patent No. 5,928,647, which are capable of inducing a class I-restricted cytotoxic T lymphocyte responses in a host.

Such compositions may also comprise buffers (*e.g.*, neutral buffered  
25 saline or phosphate buffered saline), carbohydrates (*e.g.*, glucose, mannose, sucrose or dextrans), mannitol, proteins, polypeptides or amino acids such as glycine, antioxidants, bacteriostats, chelating agents such as EDTA or glutathione, adjuvants (*e.g.*, aluminum hydroxide), solutes that render the formulation isotonic, hypotonic or weakly hypertonic with the blood of a recipient, suspending agents, thickening agents and/or preservatives.  
30 Alternatively, compositions of the present invention may be formulated as a

lyophilizate. Compounds may also be encapsulated within liposomes using well known technology.

Any of a variety of immunostimulants may be employed in the vaccines of this invention. For example, an adjuvant may be included. Most adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a stimulator of immune responses, such as lipid A, *Bordetella pertussis* or *Mycobacterium tuberculosis* derived proteins. Suitable adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, MI); Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ); AS-2 (SmithKline Beecham, Philadelphia, PA); aluminum salts such as aluminum hydroxide gel (alum) or aluminum phosphate; salts of calcium, iron or zinc; an insoluble suspension of acylated tyrosine; acylated sugars; cationically or anionically derivatized polysaccharides; polyphosphazenes; biodegradable microspheres; monophosphoryl lipid A and quil A. Cytokines, such as GM-CSF or interleukin-2, -7, or -12, may also be used as adjuvants.

Within the vaccines provided herein, the adjuvant composition is preferably designed to induce an immune response predominantly of the Th1 type. High levels of Th1-type cytokines (*e.g.*, IFN- $\gamma$ , TNF $\alpha$ , IL-2 and IL-12) tend to favor the induction of cell mediated immune responses to an administered antigen. In contrast, high levels of Th2-type cytokines (*e.g.*, IL-4, IL-5, IL-6 and IL-10) tend to favor the induction of humoral immune responses. Following application of a vaccine as provided herein, a patient will support an immune response that includes Th1- and Th2-type responses. Within a preferred embodiment, in which a response is predominantly Th1-type, the level of Th1-type cytokines will increase to a greater extent than the level of Th2-type cytokines. The levels of these cytokines may be readily assessed using standard assays. For a review of the families of cytokines, see Mosmann and Coffman, *Ann. Rev. Immunol.* 7:145-173, 1989.

Preferred adjuvants for use in eliciting a predominantly Th1-type response include, for example, a combination of monophosphoryl lipid A, preferably 3-de-O-acylated monophosphoryl lipid A (3D-MPL), together with an aluminum salt. MPL adjuvants are available from Corixa Corporation (Seattle, WA; *see* US Patent

Nos. 4,436,727; 4,877,611; 4,866,034 and 4,912,094). CpG-containing oligonucleotides (in which the CpG dinucleotide is unmethylated) also induce a predominantly Th1 response. Such oligonucleotides are well known and are described, for example, in WO 96/02555, WO 99/33488 and U.S. Patent Nos. 6,008,200 and 5,856,462. Immunostimulatory DNA sequences are also described, for example, by Sato *et al.*, *Science* 273:352, 1996. Another preferred adjuvant is a saponin, preferably QS21 (Aquila Biopharmaceuticals Inc., Framingham, MA), which may be used alone or in combination with other adjuvants. For example, an enhanced system involves the combination of a monophosphoryl lipid A and saponin derivative, such as the combination of QS21 and 3D-MPL as described in WO 94/00153, or a less reactogenic composition where the QS21 is quenched with cholesterol, as described in WO 96/33739. Other preferred formulations comprise an oil-in-water emulsion and tocopherol. A particularly potent adjuvant formulation involving QS21, 3D-MPL and tocopherol in an oil-in-water emulsion is described in WO 95/17210.

Other preferred adjuvants include Montanide ISA 720 (Seppic, France), SAF (Chiron, California, United States), ISCOMS (CSL), MF-59 (Chiron), the SBAS series of adjuvants (*e.g.*, SBAS-2 or SBAS-4, available from SmithKline Beecham, Rixensart, Belgium), Detox (Corixa, Hamilton, MT), RC-529 (Corixa, Hamilton, MT) and other aminoalkyl glucosaminide 4-phosphates (AGPs), such as those described in pending U.S. Patent Application Serial Nos. 08/853,826 and 09/074,720, the disclosures of which are incorporated herein by reference in their entireties.

Any vaccine provided herein may be prepared using well known methods that result in a combination of antigen, immune response enhancer and a suitable carrier or excipient. The compositions described herein may be administered as part of a sustained release formulation (*i.e.*, a formulation such as a capsule, sponge or gel (composed of polysaccharides, for example) that effects a slow release of compound following administration). Such formulations may generally be prepared using well known technology (*see, e.g.*, Coombes *et al.*, *Vaccine* 14:1429-1438, 1996) and administered by, for example, oral, rectal or subcutaneous implantation, or by implantation at the desired target site. Sustained-release formulations may contain a

polypeptide, polynucleotide or antibody dispersed in a carrier matrix and/or contained within a reservoir surrounded by a rate controlling membrane.

Carriers for use within such formulations are biocompatible, and may also be biodegradable; preferably the formulation provides a relatively constant level of active component release. Such carriers include microparticles of poly(lactide-co-glycolide), polyacrylate, latex, starch, cellulose, dextran and the like. Other delayed-release carriers include supramolecular biovectors, which comprise a non-liquid hydrophilic core (*e.g.*, a cross-linked polysaccharide or oligosaccharide) and, optionally, an external layer comprising an amphiphilic compound, such as a phospholipid (*see e.g.*, U.S. Patent No. 5,151,254 and PCT applications WO 94/20078, WO/94/23701 and WO 96/06638). The amount of active compound contained within a sustained release formulation depends upon the site of implantation, the rate and expected duration of release and the nature of the condition to be treated or prevented.

Any of a variety of delivery vehicles may be employed within pharmaceutical compositions and vaccines to facilitate production of an antigen-specific immune response that targets tumor cells. Delivery vehicles include antigen-presenting cells (APCs), such as dendritic cells, macrophages, B cells, monocytes and other cells that may be engineered to be efficient APCs. Such cells may, but need not, be genetically modified to increase the capacity for presenting the antigen, to improve activation and/or maintenance of the T cell response, to have anti-tumor effects *per se* and/or to be immunologically compatible with the receiver (*i.e.*, matched HLA haplotype). APCs may generally be isolated from any of a variety of biological fluids and organs, including tumor and peritumoral tissues, and may be autologous, allogeneic, syngeneic or xenogeneic cells.

Certain preferred embodiments of the present invention use dendritic cells or progenitors thereof as antigen-presenting cells. Dendritic cells are highly potent APCs (Banchereau and Steinman, *Nature* 392:245-251, 1998) and have been shown to be effective as a physiological adjuvant for eliciting prophylactic or therapeutic antitumor immunity (*see* Timmerman and Levy, *Ann. Rev. Med.* 50:507-529, 1999). In general, dendritic cells may be identified based on their typical shape (stellate *in situ*, with marked cytoplasmic processes (dendrites) visible *in vitro*), their ability to take up,



process and present antigens with high efficiency and their ability to activate naïve T cell responses. Dendritic cells may, of course, be engineered to express specific cell-surface receptors or ligands that are not commonly found on dendritic cells *in vivo* or *ex vivo*, and such modified dendritic cells are contemplated by the present invention. As  
5 an alternative to dendritic cells, secreted vesicles antigen-loaded dendritic cells (called exosomes) may be used within a vaccine (*see Zitvogel et al., Nature Med. 4:594-600, 1998*).

Dendritic cells and progenitors may be obtained from peripheral blood, bone marrow, tumor-infiltrating cells, peritumoral tissues-infiltrating cells, lymph  
10 nodes, spleen, skin, umbilical cord blood or any other suitable tissue or fluid. For example, dendritic cells may be differentiated *ex vivo* by adding a combination of cytokines such as GM-CSF, IL-4, IL-13 and/or TNF $\alpha$  to cultures of monocytes harvested from peripheral blood. Alternatively, CD34 positive cells harvested from peripheral blood, umbilical cord blood or bone marrow may be differentiated into  
15 dendritic cells by adding to the culture medium combinations of GM-CSF, IL-3, TNF $\alpha$ , CD40 ligand, LPS, flt3 ligand and/or other compound(s) that induce differentiation, maturation and proliferation of dendritic cells.

Dendritic cells are conveniently categorized as "immature" and "mature" cells, which allows a simple way to discriminate between two well characterized  
20 phenotypes. However, this nomenclature should not be construed to exclude all possible intermediate stages of differentiation. Immature dendritic cells are characterized as APC with a high capacity for antigen uptake and processing, which correlates with the high expression of Fc $\gamma$  receptor and mannose receptor. The mature phenotype is typically characterized by a lower expression of these markers, but a high  
25 expression of cell surface molecules responsible for T cell activation such as class I and class II MHC, adhesion molecules (*e.g.*, CD54 and CD11) and costimulatory molecules (*e.g.*, CD40, CD80, CD86 and 4-1BB).

APCs may generally be transfected with a polynucleotide encoding a breast tumor protein (or portion or other variant thereof) such that the breast tumor  
30 polypeptide, or an immunogenic portion thereof, is expressed on the cell surface. Such transfection may take place *ex vivo*, and a composition or vaccine comprising such

transfected cells may then be used for therapeutic purposes, as described herein. Alternatively, a gene delivery vehicle that targets a dendritic or other antigen presenting cell may be administered to a patient, resulting in transfection that occurs *in vivo*. *In vivo* and *ex vivo* transfection of dendritic cells, for example, may generally be performed using any methods known in the art, such as those described in WO 97/24447, or the gene gun approach described by Mahvi *et al.*, *Immunology and cell Biology* 75:456-460, 1997. Antigen loading of dendritic cells may be achieved by incubating dendritic cells or progenitor cells with the breast tumor polypeptide, DNA (naked or within a plasmid vector) or RNA; or with antigen-expressing recombinant bacterium or viruses (*e.g.*, vaccinia, fowlpox, adenovirus or lentivirus vectors). Prior to loading, the polypeptide may be covalently conjugated to an immunological partner that provides T cell help (*e.g.*, a carrier molecule). Alternatively, a dendritic cell may be pulsed with a non-conjugated immunological partner, separately or in the presence of the polypeptide.

Vaccines and pharmaceutical compositions may be presented in unit-dose or multi-dose containers, such as sealed ampoules or vials. Such containers are preferably hermetically sealed to preserve sterility of the formulation until use. In general, formulations may be stored as suspensions, solutions or emulsions in oily or aqueous vehicles. Alternatively, a vaccine or pharmaceutical composition may be stored in a freeze-dried condition requiring only the addition of a sterile liquid carrier immediately prior to use.

#### CANCER THERAPY

In further aspects of the present invention, the compositions described herein may be used for immunotherapy of cancer, such as breast cancer. Within such methods, pharmaceutical compositions and vaccines are typically administered to a patient. As used herein, a "patient" refers to any warm-blooded animal, preferably a human. A patient may or may not be afflicted with cancer. Accordingly, the above pharmaceutical compositions and vaccines may be used to prevent the development of a cancer or to treat a patient afflicted with a cancer. A cancer may be diagnosed using criteria generally accepted in the art, including the presence of a malignant tumor.

Pharmaceutical compositions and vaccines may be administered either prior to or following surgical removal of primary tumors and/or treatment such as administration of radiotherapy or conventional chemotherapeutic drugs. Administration may be by any suitable method, including administration by intravenous, intraperitoneal, 5 intramuscular, subcutaneous, intranasal, intradermal, anal, vaginal, topical and oral routes.

Within certain embodiments, immunotherapy may be active immunotherapy, in which treatment relies on the *in vivo* stimulation of the endogenous host immune system to react against tumors with the administration of immune 10 response-modifying agents (such as polypeptides and polynucleotides as provided herein).

Within other embodiments, immunotherapy may be passive immunotherapy, in which treatment involves the delivery of agents with established tumor-immune reactivity (such as effector cells or antibodies) that can directly or 15 indirectly mediate antitumor effects and does not necessarily depend on an intact host immune system. Examples of effector cells include T cells as discussed above, T lymphocytes (such as CD8<sup>+</sup> cytotoxic T lymphocytes and CD4<sup>+</sup> T-helper tumor-infiltrating lymphocytes), killer cells (such as Natural Killer cells and lymphokine-activated killer cells), B cells and antigen-presenting cells (such as dendritic cells and 20 macrophages) expressing a polypeptide provided herein. T cell receptors and antibody receptors specific for the polypeptides recited herein may be cloned, expressed and transferred into other vectors or effector cells for adoptive immunotherapy. The polypeptides provided herein may also be used to generate antibodies or anti-idiotypic antibodies (as described above and in U.S. Patent No. 4,918,164) for passive 25 immunotherapy.

Effector cells may generally be obtained in sufficient quantities for adoptive immunotherapy by growth *in vitro*, as described herein. Culture conditions for expanding single antigen-specific effector cells to several billion in number with retention of antigen recognition *in vivo* are well known in the art. Such *in vitro* culture 30 conditions typically use intermittent stimulation with antigen, often in the presence of cytokines (such as IL-2) and non-dividing feeder cells. As noted above,

immunoreactive polypeptides as provided herein may be used to rapidly expand antigen-specific T cell cultures in order to generate a sufficient number of cells for immunotherapy. In particular, antigen-presenting cells, such as dendritic, macrophage, monocyte, fibroblast and/or B cells, may be pulsed with immunoreactive polypeptides or transfected with one or more polynucleotides using standard techniques well known in the art. For example, antigen-presenting cells can be transfected with a polynucleotide having a promoter appropriate for increasing expression in a recombinant virus or other expression system. Cultured effector cells for use in therapy must be able to grow and distribute widely, and to survive long term *in vivo*. Studies have shown that cultured effector cells can be induced to grow in vivo and to survive long term in substantial numbers by repeated stimulation with antigen supplemented with IL-2 (see, for example, Cheever *et al.*, *Immunological Reviews* 157:177, 1997).

Alternatively, a vector expressing a polypeptide recited herein may be introduced into antigen presenting cells taken from a patient and clonally propagated *ex vivo* for transplant back into the same patient. Transfected cells may be reintroduced into the patient using any means known in the art, preferably in sterile form by intravenous, intracavitary, intraperitoneal or intratumor administration.

Routes and frequency of administration of the therapeutic compositions described herein, as well as dosage, will vary from individual to individual, and may be readily established using standard techniques. In general, the pharmaceutical compositions and vaccines may be administered by injection (*e.g.*, intracutaneous, intramuscular, intravenous or subcutaneous), intranasally (*e.g.*, by aspiration) or orally. Preferably, between 1 and 10 doses may be administered over a 52 week period. Preferably, 6 doses are administered, at intervals of 1 month, and booster vaccinations may be given periodically thereafter. Alternate protocols may be appropriate for individual patients. A suitable dose is an amount of a compound that, when administered as described above, is capable of promoting an anti-tumor immune response, and is at least 10-50% above the basal (*i.e.*, untreated) level. Such response can be monitored by measuring the anti-tumor antibodies in a patient or by vaccine-dependent generation of cytolytic effector cells capable of killing the patient's tumor cells *in vitro*. Such vaccines should also be capable of causing an immune response that

leads to an improved clinical outcome (*e.g.*, more frequent remissions, complete or partial or longer disease-free survival) in vaccinated patients as compared to non-vaccinated patients. In general, for pharmaceutical compositions and vaccines comprising one or more polypeptides, the amount of each polypeptide present in a dose  
5 ranges from about 25  $\mu$ g to 5 mg per kg of host. Suitable dose sizes will vary with the size of the patient, but will typically range from about 0.1 mL to about 5 mL.

In general, an appropriate dosage and treatment regimen provides the active compound(s) in an amount sufficient to provide therapeutic and/or prophylactic benefit. Such a response can be monitored by establishing an improved clinical  
10 outcome (*e.g.*, more frequent remissions, complete or partial, or longer disease-free survival) in treated patients as compared to non-treated patients. Increases in preexisting immune responses to a breast tumor protein generally correlate with an improved clinical outcome. Such immune responses may generally be evaluated using standard proliferation, cytotoxicity or cytokine assays, which may be performed using  
15 samples obtained from a patient before and after treatment.

#### CANCER DETECTION AND DIAGNOSIS

In general, a cancer may be detected in a patient based on the presence of one or more breast tumor proteins and/or polynucleotides encoding such proteins in a biological sample (for example, blood, sera, sputum urine and/or tumor biopsies)  
20 obtained from the patient. In other words, such proteins may be used as markers to indicate the presence or absence of a cancer such as breast cancer. In addition, such proteins may be useful for the detection of other cancers. The binding agents provided herein generally permit detection of the level of antigen that binds to the agent in the biological sample. Polynucleotide primers and probes may be used to detect the level  
25 of mRNA encoding a tumor protein, which is also indicative of the presence or absence of a cancer. In general, a breast tumor sequence should be present at a level that is at least three fold higher in tumor tissue than in normal tissue

There are a variety of assay formats known to those of ordinary skill in the art for using a binding agent to detect polypeptide markers in a sample. *See, e.g.*,  
30 Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory,

1988. In general, the presence or absence of a cancer in a patient may be determined by (a) contacting a biological sample obtained from a patient with a binding agent; (b) detecting in the sample a level of polypeptide that binds to the binding agent; and (c) comparing the level of polypeptide with a predetermined cut-off value.

5 In a preferred embodiment, the assay involves the use of binding agent immobilized on a solid support to bind to and remove the polypeptide from the remainder of the sample. The bound polypeptide may then be detected using a detection reagent that contains a reporter group and specifically binds to the binding agent/polypeptide complex. Such detection reagents may comprise, for example, a  
10 binding agent that specifically binds to the polypeptide or an antibody or other agent that specifically binds to the binding agent, such as an anti-immunoglobulin, protein G, protein A or a lectin. Alternatively, a competitive assay may be utilized, in which a polypeptide is labeled with a reporter group and allowed to bind to the immobilized binding agent after incubation of the binding agent with the sample. The extent to  
15 which components of the sample inhibit the binding of the labeled polypeptide to the binding agent is indicative of the reactivity of the sample with the immobilized binding agent. Suitable polypeptides for use within such assays include full length breast tumor proteins and portions thereof to which the binding agent binds, as described above.

The solid support may be any material known to those of ordinary skill  
20 in the art to which the tumor protein may be attached. For example, the solid support may be a test well in a microtiter plate or a nitrocellulose or other suitable membrane. Alternatively, the support may be a bead or disc, such as glass, fiberglass, latex or a plastic material such as polystyrene or polyvinylchloride. The support may also be a magnetic particle or a fiber optic sensor, such as those disclosed, for example, in U.S.  
25 Patent No. 5,359,681. The binding agent may be immobilized on the solid support using a variety of techniques known to those of skill in the art, which are amply described in the patent and scientific literature. In the context of the present invention, the term "immobilization" refers to both noncovalent association, such as adsorption, and covalent attachment (which may be a direct linkage between the agent and  
30 functional groups on the support or may be a linkage by way of a cross-linking agent). Immobilization by adsorption to a well in a microtiter plate or to a membrane is

preferred. In such cases, adsorption may be achieved by contacting the binding agent, in a suitable buffer, with the solid support for a suitable amount of time. The contact time varies with temperature, but is typically between about 1 hour and about 1 day. In general, contacting a well of a plastic microtiter plate (such as polystyrene or polyvinylchloride) with an amount of binding agent ranging from about 10 ng to about 10  $\mu$ g, and preferably about 100 ng to about 1  $\mu$ g, is sufficient to immobilize an adequate amount of binding agent.

Covalent attachment of binding agent to a solid support may generally be achieved by first reacting the support with a bifunctional reagent that will react with both the support and a functional group, such as a hydroxyl or amino group, on the binding agent. For example, the binding agent may be covalently attached to supports having an appropriate polymer coating using benzoquinone or by condensation of an aldehyde group on the support with an amine and an active hydrogen on the binding partner (*see, e.g.,* Pierce Immunotechnology Catalog and Handbook, 1991, at A12-A13).

In certain embodiments, the assay is a two-antibody sandwich assay. This assay may be performed by first contacting an antibody that has been immobilized on a solid support, commonly the well of a microtiter plate, with the sample, such that polypeptides within the sample are allowed to bind to the immobilized antibody. Unbound sample is then removed from the immobilized polypeptide-antibody complexes and a detection reagent (preferably a second antibody capable of binding to a different site on the polypeptide) containing a reporter group is added. The amount of detection reagent that remains bound to the solid support is then determined using a method appropriate for the specific reporter group.

More specifically, once the antibody is immobilized on the support as described above, the remaining protein binding sites on the support are typically blocked. Any suitable blocking agent known to those of ordinary skill in the art, such as bovine serum albumin or Tween 20™ (Sigma Chemical Co., St. Louis, MO). The immobilized antibody is then incubated with the sample, and polypeptide is allowed to bind to the antibody. The sample may be diluted with a suitable diluent, such as phosphate-buffered saline (PBS) prior to incubation. In general, an appropriate contact

time (*i.e.*, incubation time) is a period of time that is sufficient to detect the presence of polypeptide within a sample obtained from an individual with breast cancer. Preferably, the contact time is sufficient to achieve a level of binding that is at least about 95% of that achieved at equilibrium between bound and unbound polypeptide.

- 5 Those of ordinary skill in the art will recognize that the time necessary to achieve equilibrium may be readily determined by assaying the level of binding that occurs over a period of time. At room temperature, an incubation time of about 30 minutes is generally sufficient.

- 10 Unbound sample may then be removed by washing the solid support with an appropriate buffer, such as PBS containing 0.1% Tween 20™. The second antibody, which contains a reporter group, may then be added to the solid support. Preferred reporter groups include those groups recited above.

- The detection reagent is then incubated with the immobilized antibody-polypeptide complex for an amount of time sufficient to detect the bound polypeptide.
- 15 An appropriate amount of time may generally be determined by assaying the level of binding that occurs over a period of time. Unbound detection reagent is then removed and bound detection reagent is detected using the reporter group. The method employed for detecting the reporter group depends upon the nature of the reporter group. For radioactive groups, scintillation counting or autoradiographic methods are
- 20 generally appropriate. Spectroscopic methods may be used to detect dyes, luminescent groups and fluorescent groups. Biotin may be detected using avidin, coupled to a different reporter group (commonly a radioactive or fluorescent group or an enzyme). Enzyme reporter groups may generally be detected by the addition of substrate (generally for a specific period of time), followed by spectroscopic or other analysis of
- 25 the reaction products.

- To determine the presence or absence of a cancer, such as breast cancer, the signal detected from the reporter group that remains bound to the solid support is generally compared to a signal that corresponds to a predetermined cut-off value. In one preferred embodiment, the cut-off value for the detection of a cancer is the average
- 30 mean signal obtained when the immobilized antibody is incubated with samples from patients without the cancer. In general, a sample generating a signal that is three



standard deviations above the predetermined cut-off value is considered positive for the cancer. In an alternate preferred embodiment, the cut-off value is determined using a Receiver Operator Curve, according to the method of Sackett *et al.*, *Clinical Epidemiology: A Basic Science for Clinical Medicine*, Little Brown and Co., 1985, 5 p. 106-7. Briefly, in this embodiment, the cut-off value may be determined from a plot of pairs of true positive rates (*i.e.*, sensitivity) and false positive rates (100%-specificity) that correspond to each possible cut-off value for the diagnostic test result. The cut-off value on the plot that is the closest to the upper left-hand corner (*i.e.*, the value that encloses the largest area) is the most accurate cut-off value, and a sample 10 generating a signal that is higher than the cut-off value determined by this method may be considered positive. Alternatively, the cut-off value may be shifted to the left along the plot, to minimize the false positive rate, or to the right, to minimize the false negative rate. In general, a sample generating a signal that is higher than the cut-off value determined by this method is considered positive for a cancer.

15 In a related embodiment, the assay is performed in a flow-through or strip test format, wherein the binding agent is immobilized on a membrane, such as nitrocellulose. In the flow-through test, polypeptides within the sample bind to the immobilized binding agent as the sample passes through the membrane. A second, labeled binding agent then binds to the binding agent-polypeptide complex as a solution 20 containing the second binding agent flows through the membrane. The detection of bound second binding agent may then be performed as described above. In the strip test format, one end of the membrane to which binding agent is bound is immersed in a solution containing the sample. The sample migrates along the membrane through a region containing second binding agent and to the area of immobilized binding agent. 25 Concentration of second binding agent at the area of immobilized antibody indicates the presence of a cancer. Typically, the concentration of second binding agent at that site generates a pattern, such as a line, that can be read visually. The absence of such a pattern indicates a negative result. In general, the amount of binding agent immobilized on the membrane is selected to generate a visually discernible pattern when the 30 biological sample contains a level of polypeptide that would be sufficient to generate a positive signal in the two-antibody sandwich assay, in the format discussed above.

Preferred binding agents for use in such assays are antibodies and antigen-binding fragments thereof. Preferably, the amount of antibody immobilized on the membrane ranges from about 25 ng to about 1  $\mu$ g, and more preferably from about 50 ng to about 500 ng. Such tests can typically be performed with a very small amount of biological sample.

Of course, numerous other assay protocols exist that are suitable for use with the tumor proteins or binding agents of the present invention. The above descriptions are intended to be exemplary only. For example, it will be apparent to those of ordinary skill in the art that the above protocols may be readily modified to use breast tumor polypeptides to detect antibodies that bind to such polypeptides in a biological sample. The detection of such breast tumor protein specific antibodies may correlate with the presence of a cancer.

A cancer may also, or alternatively, be detected based on the presence of T cells that specifically react with a breast tumor protein in a biological sample. Within certain methods, a biological sample comprising CD4<sup>+</sup> and/or CD8<sup>+</sup> T cells isolated from a patient is incubated with a breast tumor polypeptide, a polynucleotide encoding such a polypeptide and/or an APC that expresses at least an immunogenic portion of such a polypeptide, and the presence or absence of specific activation of the T cells is detected. Suitable biological samples include, but are not limited to, isolated T cells. For example, T cells may be isolated from a patient by routine techniques (such as by Ficoll/Hypaque density gradient centrifugation of peripheral blood lymphocytes). T cells may be incubated *in vitro* for 2-9 days (typically 4 days) at 37°C with polypeptide (e.g., 5 - 25  $\mu$ g/ml). It may be desirable to incubate another aliquot of a T cell sample in the absence of breast tumor polypeptide to serve as a control. For CD4<sup>+</sup> T cells, activation is preferably detected by evaluating proliferation of the T cells. For CD8<sup>+</sup> T cells, activation is preferably detected by evaluating cytolytic activity. A level of proliferation that is at least two fold greater and/or a level of cytolytic activity that is at least 20% greater than in disease-free patients indicates the presence of a cancer in the patient.

As noted above, a cancer may also, or alternatively, be detected based on the level of mRNA encoding a breast tumor protein in a biological sample. For

example, at least two oligonucleotide primers may be employed in a polymerase chain reaction (PCR) based assay to amplify a portion of a breast tumor cDNA derived from a biological sample, wherein at least one of the oligonucleotide primers is specific for (*i.e.*, hybridizes to) a polynucleotide encoding the breast tumor protein. The amplified  
5 cDNA is then separated and detected using techniques well known in the art, such as gel electrophoresis. Similarly, oligonucleotide probes that specifically hybridize to a polynucleotide encoding a breast tumor protein may be used in a hybridization assay to detect the presence of polynucleotide encoding the tumor protein in a biological sample.

To permit hybridization under assay conditions, oligonucleotide primers  
10 and probes should comprise an oligonucleotide sequence that has at least about 60%, preferably at least about 75% and more preferably at least about 90%, identity to a portion of a polynucleotide encoding a breast tumor protein that is at least 10 nucleotides, and preferably at least 20 nucleotides, in length. Preferably, oligonucleotide primers and/or probes hybridize to a polynucleotide encoding a  
15 polypeptide described herein under moderately stringent conditions, as defined above. Oligonucleotide primers and/or probes which may be usefully employed in the diagnostic methods described herein preferably are at least 10-40 nucleotides in length. In a preferred embodiment, the oligonucleotide primers comprise at least 10 contiguous nucleotides, more preferably at least 15 contiguous nucleotides, of a DNA molecule  
20 having a sequence recited in SEQ ID NO:1-\_\_\_\_. Techniques for both PCR based assays and hybridization assays are well known in the art (*see*, for example, Mullis *et al.*, *Cold Spring Harbor Symp. Quant. Biol.*, 51:263, 1987; Erlich ed., *PCR Technology*, Stockton Press, NY, 1989).

One preferred assay employs RT-PCR, in which PCR is applied in  
25 conjunction with reverse transcription. Typically, RNA is extracted from a biological sample, such as biopsy tissue, and is reverse transcribed to produce cDNA molecules. PCR amplification using at least one specific primer generates a cDNA molecule, which may be separated and visualized using, for example, gel electrophoresis. Amplification may be performed on biological samples taken from a test patient and  
30 from an individual who is not afflicted with a cancer. The amplification reaction may be performed on several dilutions of cDNA spanning two orders of magnitude. A two-

fold or greater increase in expression in several dilutions of the test patient sample as compared to the same dilutions of the non-cancerous sample is typically considered positive.

In another embodiment, the compositions described herein may be used  
5 as markers for the progression of cancer. In this embodiment, assays as described above for the diagnosis of a cancer may be performed over time, and the change in the level of reactive polypeptide(s) or polynucleotide(s) evaluated. For example, the assays may be performed every 24-72 hours for a period of 6 months to 1 year, and thereafter performed as needed. In general, a cancer is progressing in those patients in whom the  
10 level of polypeptide or polynucleotide detected increases over time. In contrast, the cancer is not progressing when the level of reactive polypeptide or polynucleotide either remains constant or decreases with time.

Certain *in vivo* diagnostic assays may be performed directly on a tumor. One such assay involves contacting tumor cells with a binding agent. The bound  
15 binding agent may then be detected directly or indirectly via a reporter group. Such binding agents may also be used in histological applications. Alternatively, polynucleotide probes may be used within such applications.

As noted above, to improve sensitivity, multiple breast tumor protein markers may be assayed within a given sample. It will be apparent that binding agents  
20 specific for different proteins provided herein may be combined within a single assay. Further, multiple primers or probes may be used concurrently. The selection of tumor protein markers may be based on routine experiments to determine combinations that results in optimal sensitivity. In addition, or alternatively, assays for tumor proteins provided herein may be combined with assays for other known tumor antigens.

## 25 **DIAGNOSTIC KITS**

The present invention further provides kits for use within any of the above diagnostic methods. Such kits typically comprise two or more components necessary for performing a diagnostic assay. Components may be compounds, reagents, containers and/or equipment. For example, one container within a kit may  
30 contain a monoclonal antibody or fragment thereof that specifically binds to a breast

tumor protein. Such antibodies or fragments may be provided attached to a support material, as described above. One or more additional containers may enclose elements, such as reagents or buffers, to be used in the assay. Such kits may also, or alternatively, contain a detection reagent as described above that contains a reporter group suitable  
5 for direct or indirect detection of antibody binding.

Alternatively, a kit may be designed to detect the level of mRNA encoding a breast tumor protein in a biological sample. Such kits generally comprise at least one oligonucleotide probe or primer, as described above, that hybridizes to a polynucleotide encoding a breast tumor protein. Such an oligonucleotide may be used,  
10 for example, within a PCR or hybridization assay. Additional components that may be present within such kits include a second oligonucleotide and/or a diagnostic reagent or container to facilitate the detection of a polynucleotide encoding a breast tumor protein.

The following Examples are offered by way of illustration and not by way of limitation.

EXAMPLE 1

## IDENTIFICATION OF BREAST TUMOR PROTEIN CDNAS

5           This Example illustrates the identification of cDNA molecules encoding breast tumor proteins.

Potential breast-specific genes present in the GenBank human EST database were identified by electronic subtraction (similar to that described by Vasmatizis et al., *Proc. Natl. Acad. Sci. USA* 95:300-304, 1998). The sequences of EST  
10 clones (26,074) derived from various human breast cDNA libraries were obtained from the GenBank public human EST database. Each breast EST sequence was used as a query sequence in a BLASTN (National Center for Biotechnology Information) search against the human EST database. All matches considered identical (BLAST score > 40, length of matching sequence >100 base pairs, density of identical matches over this  
15 region > 80%) were grouped (aligned) together in a cluster. Clusters containing more than 200 ESTs were discarded since they probably represented repetitive elements or highly expressed genes such as those for ribosomal proteins. If two or more clusters shared common ESTs, those clusters were grouped together into a "supercluster," resulting in 9,974 breast superclusters.

20           Records for the 479 human cDNA libraries represented in the GenBank release were downloaded to create a database of these cDNA library records. These 479 cDNA libraries were grouped into three groups, Plus (normal breast and breast tumor libraries, and breast cell lines, in which expression was desired), Minus (libraries from other normal adult tissues, in which expression was not desirable), and Other  
25 (fetal tissue, infant tissue, ovary, pregnant uterus, male-specific tissues, non-breast tumors and cell lines other than breast cell lines, in which expression was considered to be irrelevant). A summary of these library groups is presented in Table I.

**TABLE I****BREAST CDNA LIBRARIES**

Library	# of Libraries
Plus	11
Normal	6
Tumor	4
Cell lines	1
Minus	171
Other	297

Each supercluster was analyzed in terms of the ESTs within the

5 supercluster. The tissue source of each EST clone was noted and used to classify the superclusters into four groups: Type 1 - EST clones derived from the Plus group libraries only; no EST clones derived from Minus or Other group libraries; Type 2 - EST clones derived from the Plus and Other group libraries only; no EST clones derived from the Minus group; Type 3 - EST clones derived from the Plus, Minus and

10 Other group libraries, but the number of EST clones derived from the Plus group is higher than from either the Minus or Other groups; and Type 4 - EST clones derived from Plus, Minus and Other group libraries, but the number of clones derived from the Plus group is higher than the number from the Minus group. For each type, subcategories were generated for clusters containing one or two clones. Types 1a, 2a,

15 3a and 4a represent Type 1, 2, 3 and 4 clusters where the number of clones in the Plus group is 2. Types 1b, 2b, 3b and 4b represent Type 1, 2, 3 and 4 clusters where the number of clones in the Plus group is 1. This analysis identified 3230 breast clusters. From these clusters, 2501 EST clones were ordered from Research Genetics, Inc. (Huntsville, AL), and were received as frozen glycerol stocks in 96-well plates. A

20 summary of the clusters generated and clones ordered is shown in Table II.

**TABLE II****BREAST CLUSTER SUMMARY**

Type	# of Superclusters	# of ESTs Ordered
1	50	49
1a (2/0/0)	130	130
1b (1/0/0)	65	65
2	167	162
2a (2/0/n)	312	321
2a (1/0/n)	1875	1752
3	56	22
3a	32	0
4	333	0
4a	210	0
Total	3230	2501

The EST clone inserts were PCR-amplified for Synteni microarray  
 5 analysis using amino-linked PCR primers. When more than one PCR product was  
 obtained for a particular clone, that clone was not used for expression analysis. In total,  
 1896 clones from the electronic subtraction method were analyzed by microarray  
 analysis to identify electronic subtraction breast clones that had high tumor vs. normal  
 tissue mRNA. Such screens were performed using a Synteni (Palo Alto, CA)  
 10 microarray, according to the manufacturer's instructions (and essentially as described by  
 Schena et al., *Proc. Natl. Acad. Sci. USA* 93:10614-10619, 1996 and Heller et al., *Proc.*  
*Natl. Acad. Sci. USA* 94:2150-2155, 1997). Within these analyses, the clones were  
 arrayed on the chip, which was then probed with fluorescent probes generated from  
 normal and tumor breast cDNA, as well as various other normal tissues. The slides  
 15 were scanned and the fluorescence intensity was measured.

Clones with an expression ratio greater than 2 (*i.e.*, the level in breast  
 tumor cDNA was at least twice the level in normal breast cDNA) were identified as



breast tumor-specific sequences (Table III). The sequences of these clones are provided in SEQ ID NOs:1-124.

**TABLE III**

5

**BREAST-TUMOR SPECIFIC CLONES**

SEQ ID NO.	Sequence Designation
1	19702
2	19703
3	B1003C
4	B1002C
5	19708
6	19709
7	19710
8	19711
9	B1006C
10	B1007C
11	19714
12	19715
13	19716
14	19717
15	19718
16	19719
17	19720
18	19721
19	19997
20	19998
21	19999
22	20000
23	20001
24	20002
25	20005
26	20006
27	20007
28	20008
29	20009
30	20010
31	20011
32	20012
33	20013
34	20014

35	20079
36	20080
37	20081
38	20082
39	20083
40	20085
41	20086
42	20087
43	20088
44	20089
45	20090
46	20091
47	20092
48	20093
49	20115
50	20116
51	20118
52	20119
53	20120
54	20121
55	20122
56	20123
57	20124
58	20125
59	20360
60	20361
61	20362
62	22180
63	22181
64	22182
65	22183
66	22185
67	22186
68	22188
69	22189
70	22190
71	22191
72	22192
73	22193
74	22194
75	22196
76	22197
77	22198
78	22199
79	22200
80	22201

81	22202
82	22204
83	22206
84	22207
85	22208
86	22209
87	22210
88	22211
89	22212
90	22213
91	22214
92	22215
93	22216
94	22217
95	22218
96	22219
97	22220
98	22221
99	22222
100	22223
101	22224
102	22225
103	22226
104	22227
105	22228
106	22229
107	22230
108	22231
109	22232
110	22233
111	22234
112	22235
113	22236
114	22237
115	22238
116	22239
117	22240
118	22241
119	22242
120	22243
121	22244
122	22245
123	22334
124	22335

The B1002C sequence (SEQ ID NO:4; 517 bp) was used in a BlastN search of the GenBank Human EST database to identify overlapping sequences that extended further toward the 5' end of the corresponding gene. Two human EST clones were identified in this manner and were purchased from Genome Systems (St. Louis, MO) and sequenced. The resulting sequence information gave rise to a partial predicted open reading frame of 214 amino acids, which is 88% identical over 180 amino acids to the mouse "iroquois homeobox protein 3" (GenBank protein accession #Y15001). The extended B1002C sequence, along with the predicted open reading frame, are provided herein as SEQ ID NOs:125 and 126, respectively. The alignment between B1002C and the mouse iroquois homeobox protein 3 is presented in Figure 1.

## EXAMPLE 2

### SYNTHESIS OF POLYPEPTIDES

15

Polypeptides may be synthesized on a Perkin Elmer/Applied Biosystems Division 430A peptide synthesizer using Fmoc chemistry with HPTU (O-Benzotriazole-N,N,N',N'-tetramethyluronium hexafluorophosphate) activation. A Gly-Cys-Gly sequence may be attached to the amino terminus of the peptide to provide a method of conjugation, binding to an immobilized surface, or labeling of the peptide. Cleavage of the peptides from the solid support may be carried out using the following cleavage mixture: trifluoroacetic acid:ethanedithiol:thioanisole:water:phenol (40:1:2:2:3). After cleaving for 2 hours, the peptides may be precipitated in cold methyl-t-butyl-ether. The peptide pellets may then be dissolved in water containing 0.1% trifluoroacetic acid (TFA) and lyophilized prior to purification by C18 reverse phase HPLC. A gradient of 0%-60% acetonitrile (containing 0.1% TFA) in water (containing 0.1% TFA) may be used to elute the peptides. Following lyophilization of the pure fractions, the peptides may be characterized using electrospray or other types of mass spectrometry and by amino acid analysis.

30

From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

## CLAIMS

What is claimed:

1. An isolated polypeptide, comprising at least an immunogenic portion of a breast tumor protein, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide sequence selected from the group consisting of:
  - (a) sequences recited in SEQ ID NOs:1, 3, 4, 8, 9, 11, 12, 14, 15, 17, 26, 30, 31, 35, 36, 47, 52, 55, 60, 62, 63, 66, 69, 71, 72, 74-78, 81, 83-87, 92, 94, 97, 99-104, 107, 108, 110, 112, 113, 117, 118, 120-122 or 125;
  - (b) sequences that hybridize to a sequence recited in any one of SEQ ID NOs:1, 3, 4, 8, 9, 11, 12, 14, 15, 17, 26, 30, 31, 35, 36, 47, 52, 55, 60, 62, 63, 66, 69, 71, 72, 74-78, 81, 83-87, 92, 94, 97, 99-104, 107, 108, 110, 112, 113, 117, 118, 120-122 or 125, under moderately stringent conditions; and
  - (c) complements of sequences of (a) or (b).
2. An isolated polypeptide according to claim 1, wherein the polypeptide comprises an amino acid sequence that is encoded by a polynucleotide sequence recited in any one of SEQ ID NOs:1, 3, 4, 8, 9, 11, 12, 14, 15, 17, 26, 30, 31, 35, 36, 47, 52, 55, 60, 62, 63, 66, 69, 71, 72, 74-78, 81, 83-87, 92, 94, 97, 99-104, 107, 108, 110, 112, 113, 117, 118, 120-122 or 125, or a complement of any of the foregoing polynucleotide sequences.
3. An isolated polypeptide comprising a sequence recited in SEQ ID NO:126.
4. An isolated polynucleotide encoding at least 15 amino acid residues of a breast tumor protein, or a variant thereof that differs in one or more substitutions, deletions, additions and/or insertions such that the ability of the variant to react with antigen-specific antisera is not substantially diminished, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide comprising a sequence recited in any one of SEQ

ID NOs:1, 3, 4, 8, 9, 11, 12, 14, 15, 17, 26, 30, 31, 35, 36, 47, 52, 55, 60, 62, 63, 66, 69, 71, 72, 74-78, 81, 83-87, 92, 94, 97, 99-104, 107, 108, 110, 112, 113, 117, 118, 120-122 or 125, or a complement of any of the foregoing sequences.

5. An isolated polynucleotide encoding a breast tumor protein, or a variant thereof, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide comprising a sequence recited in any one of SEQ ID NOs:1, 3, 4, 8, 9, 11, 12, 14, 15, 17, 26, 30, 31, 35, 36, 47, 52, 55, 60, 62, 63, 66, 69, 71, 72, 74-78, 81, 83-87, 92, 94, 97, 99-104, 107, 108, 110, 112, 113, 117, 118, 120-122 or 125, or a complement of any of the foregoing sequences.

6. An isolated polynucleotide, comprising a sequence recited in any one of SEQ ID NOs: 1, 3, 4, 8, 9, 11, 12, 14, 15, 17, 26, 30, 31, 35, 36, 47, 52, 55, 60, 62, 63, 66, 69, 71, 72, 74-78, 81, 83-87, 92, 94, 97, 99-104, 107, 108, 110, 112, 113, 117, 118, 120-122 or 125.

7. An isolated polynucleotide, comprising a sequence that hybridizes to a sequence recited in any one of SEQ ID NOs:1, 3, 4, 8, 9, 11, 12, 14, 15, 17, 26, 30, 31, 35, 36, 47, 52, 55, 60, 62, 63, 66, 69, 71, 72, 74-78, 81, 83-87, 92, 94, 97, 99-104, 107, 108, 110, 112, 113, 117, 118, 120-122 or 125, under moderately stringent conditions.

8. An isolated polynucleotide complementary to a polynucleotide according to any one of claims 4-7.

9. An expression vector, comprising a polynucleotide according to any one of claims claim 4-8.

10. A host cell transformed or transfected with an expression vector according to claim 9.

11. An isolated antibody, or antigen-binding fragment thereof, that specifically binds to a breast tumor protein that comprises an amino acid sequence that is encoded by a polynucleotide sequence recited in any one of SEQ ID NOs:1, 3, 4, 8, 9, 11, 12, 14, 15, 17, 26, 30, 31, 35, 36, 47, 52, 55, 60, 62, 63, 66, 69, 71, 72, 74-78, 81, 83-87, 92, 94, 97, 99-104, 107, 108, 110, 112, 113, 117, 118, 120-122 or 125, or a complement of any of the foregoing polynucleotide sequences.

12. A fusion protein, comprising at least one polypeptide according to claim 1.

13. A fusion protein according to claim 12, wherein the fusion protein comprises an expression enhancer that increases expression of the fusion protein in a host cell transfected with a polynucleotide encoding the fusion protein.

14. A fusion protein according to claim 12, wherein the fusion protein comprises a T helper epitope that is not present within the polypeptide of claim 1.

15. A fusion protein according to claim 12, wherein the fusion protein comprises an affinity tag.

16. An isolated polynucleotide encoding a fusion protein according to claim 12.

17. A pharmaceutical composition, comprising a physiologically acceptable carrier and at least one component selected from the group consisting of:

- (a) a polypeptide according to claim 1;
- (b) a polynucleotide according to claim 4;
- (c) an antibody according to claim 11;
- (d) a fusion protein according to claim 12; and
- (e) a polynucleotide according to claim 16.



18. A vaccine comprising an immunostimulant and at least one component selected from the group consisting of:

- (a) a polypeptide according to claim 1;
- (b) a polynucleotide according to claim 4;
- (c) an antibody according to claim 11;
- (d) a fusion protein according to claim 12; and
- (e) a polynucleotide according to claim 16.

19. A vaccine according to claim 18, wherein the immunostimulant is an adjuvant.

20. A vaccine according to any claim 18, wherein the immunostimulant induces a predominantly Type I response.

21. A method for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of a pharmaceutical composition according to claim 17.

22. A method for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of a vaccine according to claim 18.

23. A pharmaceutical composition comprising an antigen-presenting cell that expresses a polypeptide according to claim 1, in combination with a pharmaceutically acceptable carrier or excipient.

24. A pharmaceutical composition according to claim 23, wherein the antigen presenting cell is a dendritic cell or a macrophage.

25. A vaccine comprising an antigen-presenting cell that expresses a polypeptide comprising at least an immunogenic portion of a breast tumor protein, or a variant thereof, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide sequence selected from the group consisting of:

- (a) sequences recited in SEQ ID NOs:1-125;
- (b) sequences that hybridize to a sequence recited in any one of SEQ ID NOs:1-125 under moderately stringent conditions; and
- (c) complements of sequences of (i) or (ii);  
in combination with an immunostimulant.

26. A vaccine according to claim 25, wherein the immunostimulant is an adjuvant.

27. A vaccine according to claim 25, wherein the immunostimulant induces a predominantly Type I response.

28. A vaccine according to claim 25, wherein the antigen-presenting cell is a dendritic cell.

29. A method for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of an antigen-presenting cell that expresses a polypeptide comprising at least an immunogenic portion of a breast tumor protein, or a variant thereof, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide sequence selected from the group consisting of:

- (a) sequences recited in SEQ ID NOs:1-125;
- (b) sequences that hybridize to a sequence recited in any one of SEQ ID NOs:1-125 under moderately stringent conditions; and
- (c) complements of sequences of (i) or (ii) encoded by a polynucleotide recited in any one of SEQ ID NOs:1-125;

and thereby inhibiting the development of a cancer in the patient.

30. A method according to claim 29, wherein the antigen-presenting cell is a dendritic cell.

31. A method according to any one of claims 21, 22 and 29, wherein the cancer is breast cancer.

32. A method for removing tumor cells from a biological sample, comprising contacting a biological sample with T cells that specifically react with a breast tumor protein, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide sequence selected from the group consisting of:

- (i) polynucleotides recited in any one of SEQ ID NOs:1-125; and
- (ii) complements of the foregoing polynucleotides;

wherein the step of contacting is performed under conditions and for a time sufficient to permit the removal of cells expressing the antigen from the sample.

33. A method according to claim 32, wherein the biological sample is blood or a fraction thereof.

34. A method for inhibiting the development of a cancer in a patient, comprising administering to a patient a biological sample treated according to the method of claim 32.

35. A method for stimulating and/or expanding T cells specific for a breast tumor protein, comprising contacting T cells with at least one component selected from the group consisting of:

(a) polypeptides comprising at least an immunogenic portion of a breast tumor protein, or a variant thereof, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide sequence selected from the group consisting of:

(i) sequences recited in SEQ ID NOs:1-125;

(ii) sequences that hybridize to a sequence recited in any one of SEQ ID NOs:1-125 under moderately stringent conditions; and

(iii) complements of sequences of (i) or (ii);

(b) polynucleotides encoding a polypeptide of (a); and

(c) antigen presenting cells that express a polypeptide of (a);

under conditions and for a time sufficient to permit the stimulation and/or expansion of T cells.

36. An isolated T cell population, comprising T cells prepared according to the method of claim 35.

37. A method for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of a T cell population according to claim 36.

38. A method for inhibiting the development of a cancer in a patient, comprising the steps of:

(a) incubating CD4<sup>+</sup> and/or CD8<sup>+</sup> T cells isolated from a patient with at least one component selected from the group consisting of:

(i) polypeptides comprising at least an immunogenic portion of a breast tumor protein, or a variant thereof, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide sequence selected from the group consisting of:

(1) sequences recited in SEQ ID NOs:1-125;

(2) sequences that hybridize to a sequence recited in any one of SEQ ID NOs:1-125 under moderately stringent conditions; and

- (3) complements of sequences of (1) or (2);
  - (ii) polynucleotides encoding a polypeptide of (i); and
  - (iii) antigen presenting cells that expresses a polypeptide of (i);
- such that T cells proliferate; and
- (b) administering to the patient an effective amount of the proliferated T cells,
- and thereby inhibiting the development of a cancer in the patient.

39. A method for inhibiting the development of a cancer in a patient, comprising the steps of:

- (a) incubating CD4<sup>+</sup> and/or CD8<sup>+</sup> T cells isolated from a patient with at least one component selected from the group consisting of:

- (i) polypeptides comprising at least an immunogenic portion of a breast tumor protein, or a variant thereof, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide sequence selected from the group consisting of:

- (1) sequences recited in SEQ ID NOs:1-125;
    - (2) sequences that hybridize to a sequence recited in any one of SEQ ID NOs:1-125 under moderately stringent conditions; and

- (3) complements of sequences of (1) or (2);
      - (ii) polynucleotides encoding a polypeptide of (i); and
      - (iii) antigen presenting cells that express a polypeptide of (i);
- such that T cells proliferate;
- (b) cloning at least one proliferated cell to provide cloned T cells; and
  - (c) administering to the patient an effective amount of the cloned T cells, and
- thereby inhibiting the development of a cancer in the patient.

40. A method for determining the presence or absence of a cancer in a patient, comprising the steps of:

- (a) contacting a biological sample obtained from a patient with a binding agent that binds to a breast tumor protein, wherein the tumor protein comprises an amino acid

sequence that is encoded by a polynucleotide sequence recited in any one of SEQ ID NOs:1-125 or a complement of any of the foregoing polynucleotide sequences;

(b) detecting in the sample an amount of polypeptide that binds to the binding agent; and

(c) comparing the amount of polypeptide to a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient.

41. A method according to claim 40, wherein the binding agent is an antibody.

42. A method according to claim 43, wherein the antibody is a monoclonal antibody.

43. A method according to claim 40, wherein the cancer is breast cancer.

44. A method for monitoring the progression of a cancer in a patient, comprising the steps of:

(a) contacting a biological sample obtained from a patient at a first point in time with a binding agent that binds to a breast tumor protein, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide sequence recited in any one of SEQ ID NOs:1-125 or a complement of any of the foregoing polynucleotide sequences;

(b) detecting in the sample an amount of polypeptide that binds to the binding agent;

(c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and

(d) comparing the amount of polypeptide detected in step (c) to the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

45. A method according to claim 44, wherein the binding agent is an antibody.

46. A method according to claim 45, wherein the antibody is a monoclonal antibody.
47. A method according to claim 44, wherein the cancer is a breast cancer.
48. A method for determining the presence or absence of a cancer in a patient, comprising the steps of:
- (a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a breast tumor protein, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide sequence recited in any one of SEQ ID NO:1-125 or a complement of any of the foregoing polynucleotide sequences;
  - (b) detecting in the sample an amount of a polynucleotide that hybridizes to the oligonucleotide; and
  - (c) comparing the amount of polynucleotide that hybridizes to the oligonucleotide to a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient.
49. A method according to claim 48, wherein the amount of polynucleotide that hybridizes to the oligonucleotide is determined using a polymerase chain reaction.
50. A method according to claim 48, wherein the amount of polynucleotide that hybridizes to the oligonucleotide is determined using a hybridization assay.
51. A method for monitoring the progression of a cancer in a patient, comprising the steps of:
- (a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a breast tumor protein, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide

sequence recited in any one of SEQ ID NO:1-125 or a complement of any of the foregoing polynucleotide sequences;

(b) detecting in the sample an amount of a polynucleotide that hybridizes to the oligonucleotide;

(c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and

(d) comparing the amount of polynucleotide detected in step (c) to the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

52. A method according to claim 51, wherein the amount of polynucleotide that hybridizes to the oligonucleotide is determined using a polymerase chain reaction.

53. A method according to claim 51, wherein the amount of polynucleotide that hybridizes to the oligonucleotide is determined using a hybridization assay.

54. A diagnostic kit, comprising:

- (a) one or more antibodies according to claim 11; and
- (b) a detection reagent comprising a reporter group.

55. A kit according to claim 54, wherein the antibodies are immobilized on a solid support.

56. A kit according to claim 54, wherein the detection reagent comprises an anti-immunoglobulin, protein G, protein A or lectin.

57. A kit according to claim 54, wherein the reporter group is selected from the group consisting of radioisotopes, fluorescent groups, luminescent groups, enzymes, biotin and dye particles.



58. An oligonucleotide comprising 10 to 40 contiguous nucleotides that hybridize under moderately stringent conditions to a polynucleotide that encodes a breast tumor protein, wherein the tumor protein comprises an amino acid sequence that is encoded by a polynucleotide sequence recited in any one of SEQ ID NOs:1, 3, 4, 8, 9, 11, 12, 14, 15, 17, 26, 30, 31, 35, 36, 47, 52, 55, 60, 62, 63, 66, 69, 71, 72, 74-78, 81, 83-87, 92, 94, 97, 99-104, 107, 108, 110, 112, 113, 117, 118, 120-122 or 125, or a complement of any of the foregoing polynucleotides.

59. A oligonucleotide according to claim 58, wherein the oligonucleotide comprises 10-40 contiguous nucleotides recited in any one of SEQ ID NOs:1, 3, 4, 8, 9, 11, 12, 14, 15, 17, 26, 30, 31, 35, 36, 47, 52, 55, 60, 62, 63, 66, 69, 71, 72, 74-78, 81, 83-87, 92, 94, 97, 99-104, 107, 108, 110, 112, 113, 117, 118, 120-122 or 125.

60. A diagnostic kit, comprising:

- (a) an oligonucleotide according to claim 59; and
- (b) a diagnostic reagent for use in a polymerase chain reaction or hybridization assay.

1/1

1  
1 H S F P Q L G Y Q Y I R P L Y P P E R F A A A G G G G S S A G R S G F G A G A S E L A A S G S R1002Corf. PRO  
Iroquois. PRO  
1  
51 L S N V L S S V Y G A P Y A A A A A A A A A A Q G Y G A F L P Y A T E L P F F Q L G A Q Y E L K D R1002Corf. PRO  
Iroquois. PRO  
1  
101 S P G V Q H P A T A A A F P H P A F Y P Y G Q Y Q F G D F S R F K H A T R E S T S T L K A W L N R1002Corf. PRO  
Iroquois. PRO  
1  
151 E H R K N P Y P T K G E K I M L A I I T K M T L T Q V S T W F A A R R R L K K E H K H T W A P R S R1002Corf. PRO  
Iroquois. PRO  
1  
201 R T D E R G N A Y G S E R E E E D E E E D F E E S K R E L F E F E E F L A G R G G G I G G R Q A G R R1002Corf. PRO  
Iroquois. PRO  
1  
251 R R R D E E I D L E N L D S A A G S E L I L A G A A H R N G D F G L G P I S D C K T S D S D S S R1002Corf. PRO  
Iroquois. PRO  
31 F E R G A G R G S S G O R E E G A R R R E F G L A P R R V V U Q S L U P C A P A F A F A S A L O K F R1002Corf. PRO  
301 E G L E D R P L S V L S L A P P P P V A R A P A S P F S F S S L D P C A P A F A F S A L O K F Iroquois. PRO  
61 K I M S L A R T A T S P D N P R R S P P G A H G S F F G A A V A F S A L O S F A A A A A A A A R1002Corf. PRO  
351 K I M S L A E T A T S P D N P R R S P P G A G S F P G A A V A P F L Q L S F A A A A A A A A A I I Iroquois. PRO  
109 R I V S A P L O K F P A M T N R P F P G P F F G F R F H P L S L G S A F F H L L G L F G A A G H F R1002Corf. PRO  
401 R I V S A P L O K F P A M T N R P F P G P F G R P H P L S H L G S A P O H L L G L F G A A G H F Iroquois. PRO  
159 A A A A A - F A R T A F F G G T D R C S A L E V E K K L L K T A F O V P P P P P O H L D A A L V R1002Corf. PRO  
451 A A A A A Y A R P A E P E S O T D R C S A L E V E K K L L K T A F O V P P P P P O H L D A A L V Iroquois. PRO  
200 L S A L R R R R  
501 L S A L R R R R

FIG 1

## SEQUENCE LISTING

<110> Corixa Corporation  
 Mitcham, Jennifer L.  
 Jiang, Yuqiu

<120> COMPOSITIONS AND METHODS FOR THE THERAPY  
 AND DIAGNOSIS OF BREAST CANCER

<130> 210121.482PC

<140> PCT

<141> 2000-06-23

<160> 127

<170> FastSEQ for Windows Version 3.0

<210> 1

<211> 476

<212> DNA

<213> Homo sapien

<400> 1

tgaagaaatg ggtctcttgc cccctcccag taaaagaata ttttttttaa aaaaaaaggc	60
tctgctttga cctgaagtat tttatctatc ctgagtctca ggacactggt gatggaatta	120
aggccaagca catctgcaaa aaagacattg ctggaggagg tgcaaagagc tggaaaccaa	180
gtctccagtc ctgggaaaag cagtggatg gaaaagcaat ggaaagagca ttttgaaaat	240
gccattccac tgttttctgg cctttatgat ttctgctgag aaatccactg ttagtctgat	300
ggggtctccc ttcatagcac caatgacctg aagagccttg ttgaaggaag actccatctg	360
atgactcaga gcaagtattt tttagtgtgt tattgttatt agcaagaaag agggccataa	420
aatacatggg gcaagctgaa tatatcttag gcaaaaaraag ggaatattca aattct	476

<210> 2

<211> 680

<212> DNA

<213> Homo sapien

<220>

<221> misc\_feature

<222> (1)...(680)

<223> n = A,T,C or G

<400> 2

acttgccctc gaggccaaga attcggcacn aggatccttc cagctgggtct gctctgagtg	60
gcttatcccg catgagcagg agcgtgctgg ccctgagtac tgaactttct gagtaacaat	120
gagacacggt acagaacctg tgttcagggt gcgggtgagc tgccctctcc aaatccagcc	180
agagatgcac attcctcggc cagtctcagc caacagtacc aaaagtgatt tttgagtgtg	240
ccagggtaaa ggcttcagtc tcagcctcag ttattttaga caatctcgcc atctttaatt	300
tcttagcttc ctgttctaatt aaatgcacgg ctttaccttt cctgtcagaa ataaaccaag	360
gctctaaaag atgatttccc ttctgtaact ccctagagcc acagggtctc attccttttc	420
ccattatact tctcacaatt cagtttctat gagtttgatc acctgatttt ttaacaaaa	480
tatttctaac gggaatgggt gggagtgtgt gtgaaaagag gtgaaatgtg gttgtatgag	540

```

ccaatcatat ttgtgatttt ttaaaaaaag tttaaaagga aatatctgtt ctgaaacccc 600
acttaagcat tgtttttata taaaaacaat gataaaaatg tgaaactgtg aaataaatat 660
accatattag cgtaccacc 680

```

```

<210> 3
<211> 474
<212> DNA
<213> Homo sapien

```

```

<400> 3
ggcacgagga tgatggaaga agaagaagga ggaggaggag aaggagcaag aggaagagga 60
agaagaagaa caagaagaga agaagaagaa gaggaagaag aggaagagga agaagaagaa 120
agaaagaaga aagaacaaga agaagaagga agaaaagaag tgggaagaaa gcatgaatag 180
gagtagcttc acaaagatag gagataaccg agaaactgtt tttattgcat cttcctgtgt 240
aggccacctt gtaagctgtg gcctcagact taccttttat cttgtgtaat gcttttttac 300
cactattgtg tctgctgctt atgtcttttc caggcatgta ctgttgcaagg ggtgtccaat 360
catttagctt ttctgagaca cactggataa acacaaattg tcctgggcca cactgaaaat 420
acattaacac taacaatagc gtgataagct caaattaaaa aaaaaaaaaa aaaa 474

```

```

<210> 4
<211> 517
<212> DNA
<213> Homo sapien

```

```

<400> 4
gtagtgctt ggaagtggag aaaaagttac tcaagacagc tttccagccc gtgcccaggc 60
ggccccagaa ccatygtggac gccgccctgg tcttatcggc tctctcctca tcctagttct 120
ttaaaaaaaa caaaaaaaca aaaaaaactt tttttaatcg ttgtaataat tgtataaaaa 180
aaatcgctct gtatagttac aacttgtaag catgtccgtg tataaataacc taaaagcaaa 240
actaaacaaa gaaagtaaga aaaagaaata aaaccagtcc tcctcagccc tccccagtc 300
gcttctgtgg caccctgcct tcgtgtgag gttgtttgt cgggttgatt ttggggggtg 360
gagtttcagt gagaataaac gtgtctgcct ttgtgtgtgt gtatatatac agagaaatgt 420
acatatgtgt gaaccaaatt gtacgagaaa gtatctattt ttggctaaat aaatgagctg 480
cctgccactt tgactataaa aaaaaaaaaa aaaaaaa 517

```

```

<210> 5
<211> 499
<212> DNA
<213> Homo sapien

```

```

<400> 5
ctagaaggag aaatactttc tgtcttatta aaattaataa attattggtc tttacaagac 60
ttggatacat tacagcagac atggaaatat aattttaaaa aatttctctc caacctcctt 120
caaattcagt caccactgtt atattacctt ctccaggaac cctccagtgg ggaaggctgc 180
gatattagat ttccttgat gcaaagtttt tgttgaaagc tgtgctcaga ggaggtgaga 240
ggagaggaag gagaaaactg catcataact ttacagaatt gaatctagag tcttccccga 300
aaagcccaga aacttctctg cagtatctgg cttgtccatc tgggtctaagg tggctgcttc 360
ttccccagcc atgagtcagt ttgtgcccac gaataataca cgacctgtta tttccatgac 420
tgctttactg tatttttaag gtcaatatata tgtacatttg ataataaaat aatattctcc 480
caaaaaaaaa aaaaaaaaaa 499

```

```

<210> 6
<211> 88
<212> DNA
<213> Homo sapien

```

<220>  
 <221> misc\_feature  
 <222> (1)...(88)  
 <223> n = A,T,C or G

<400> 6  
 ttacccccca aagctgacgc ctcatttgca ctgtcnagcc ttgaactctt ctgccagccg 60  
 tatcttatat acatggtatc catgttgc 88

<210> 7  
 <211> .777  
 <212> DNA  
 <213> Homo sapien

<400> 7  
 aasttgacct gcctgggtcac agacctgacc acctatgaca gcgtgaccat cctggacccg 60  
 ccagaatggc gaagctgtga aaaccacac caacatctcc gagagccacc ccaatgccac 120  
 tttcagcgcc gtgggtgagg ccagcatctg cgaggatgac tgggaattccg gggagagggtt 180  
 cacgtgcacc gtgaccacaca cagacctgcc ctgccactg aagcagacca tctcccgcc 240  
 caaggggggtg gccctgcaca ggcccgatgt ctacttgctg ccaccagccc gggagcagct 300  
 gaacctgcgg gagtcggcca ccatcacgtg cctggtgacg ggcttctctc ccgcggaagt 360  
 ctctgtgcag tggatgcaga gggggcagcc cttgtccccg gagaagtatg tgaccagcgc 420  
 cccaatgcct gagccccagg cccagggccg gtacttcgcc cacagcatcc tgaccgtgtc 480  
 cgaaagaggaa tggaaacacgg gggagaccta cacctgcgtg gtggcccatg aggccctgcc 540  
 caacagggtc accgagagga ccgtggacaa gtccaccggt aaactcacc tgtacaacgt 600  
 gtccctgggtc atgtccgaca cagctggcac ctgctactga ccctgctggc ctgccacag 660  
 gctcggggcg gctggccgct ctgtgtgtgc atgcaacta accgtgtcaa cggggtaaga 720  
 tgttgcatct tataaaatta gaaataaaaa gatccattca aaaaaaaaaa aaaaaaa 777

<210> 8  
 <211> 996  
 <212> DNA  
 <213> Homo sapien

<400> 8  
 gccctcgggc caagaattcg gcacgagggc tactggatgc actgtcagat taacttattt 60  
 cattaagaa gttgggagaa caaataggaa aaaaaaact tatttttcta gtaaatatta 120  
 atgtattaca tttcaaataa tgggtgcctga catattgaat aattattttc tacagtgtac 180  
 gtatgcaaca aagatattcc atcatgcatt agagtcagtt ctggctctgc ctagtgttt 240  
 acatttgcaa atgtagcaaa caaggtaatg aagcaactat ttctattgca gtagatatcc 300  
 ttttgtgtgt gtgtgtgtgc attaaagttg taaacggtaa catgaaacaa atgaaagttc 360  
 ttgtatataat ggtatggaaa acaagaagga aatgaaaata tttttatgcc tacttaggaa 420  
 aaaaagggtg gcacttattc attccaagta cttttttttt ttaattttta agctcttaac 480  
 tcacattggt atgcttaaga tgataaacat atatcctctt tttattgctt tgtctatgtt 540  
 tcatatgaaa catttcagaa attattttga taagtgttgc tggaaatctgc aacgctgatg 600  
 ttttttttgc attctgtagt cgcatttgca ctccattttt acattaattc gcagttgctt 660  
 tgtatcattg ttttgtttgg gttttgtttc tttttcacag tgccgggtct tcgtttctta 720  
 aagttggatg gcaggtagag ttcaaccagt tcgtgactgt tgtagcgaat gaagttaaaa 780  
 aaatgtcttt ctgatgttgt gttgtcattt tcatttttgc atttttttgt ttgcatatta 840  
 aaaaaagaga aaagagaaag caagagacag aaatcaggac taagtccctc gcttcagttt 900  
 cattgttaac gggccttatt ctgatctcac ctgtcgcgta gctctaatat tcacataaac 960  
 tgaaataaag aagtggaatg aggagctttg acagtc 996

<210> 9

<211> 811  
 <212> DNA  
 <213> Homo sapien

<400> 9

ggcacgaggc	tgactttgat	ggctccacaa	acataaccca	gtgtaaaaaac	agaagatgtg	60
gaggggagct	gggagatttc	actgggtaca	ttgaatcccc	aaactaccca	ggcaattacc	120
cagccaacac	cgagtgtacg	tggaccatca	acccaccccc	caagcgccgc	atcctgatcg	180
tgggccctga	gatcttcctg	cccatagagg	acgactgtgg	ggactatctg	gtgatgcgga	240
aaacctcttc	atccaattct	gtgacaacat	atgaaacctg	ccagacctac	gaacgcccc	300
tgccttcac	ctccagggtca	aagaagctgt	ggattcagtt	caagtccaat	gaaggggaaca	360
gcgctagagg	gttccaggtc	ccatacgtga	catatgatga	ggactaccag	gaactcattg	420
aagacatagt	tcgagatggc	aggctctatg	catctgagaa	ccatcaggaa	atacttaagg	480
ataagaaact	tatcaaggct	ctgtttgatg	tcctggcccc	tccccagAAC	tatttcaagt	540
acacagccca	ggagtcccg	gagatgttcc	caagatcggt	catccgattg	ctacgttcca	600
aagtgtccag	gtttttgaga	ctttacaaat	gactcagccc	acgtgccact	caatacaaat	660
gttctgctat	agggttgggtg	ggacagagct	gtcttctctc	tgcattgtcag	cacagtcggg	720
tattgtctgc	tcccgatca	gtgactcatt	agagttcaat	ttttatagat	aattcagata	780
ttttggtaaa	ttgaaaaaaa	aaaaaaaaaa	a			811

<210> 10  
 <211> 542  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(542)  
 <223> n = A,T,C or G

<400> 10

tngccctcga	ggccaagaat	tcggcacgag	ggttttctgt	actttgccta	cggcagcaac	60
ctgctgacag	agaggatcca	cctccgaaac	ccctcggcgg	cgtnntctg	tgtggccgc	120
ctgcaggcaa	gaaggggtta	aaagtggat	gtatgttgta	atagaagtta	aagttgcaac	180
tcaagaagga	aaagaaataa	cctgtcgaag	ttatctgatg	acaaattacg	aaagtgtccc	240
cccatcccca	cagtataaaa	agattatttg	catgggtgca	aaagaaaatg	gtttgccgct	300
ggagtatcaa	gagaagttaa	aagcaataga	accaaataac	tatacaggaa	aggtctcaga	360
agaaattgaa	gacatcatca	aaaaggggga	aacacaaact	ctttagaaca	taacagaata	420
tatctaaggg	tattctatgt	gctaataata	aatattttta	acacttgaga	acagggatct	480
gggggatctc	cacgtttgat	ccattttcag	cagtgtctctg	aaggagtatc	ttacttgggt	540
ga						542

<210> 11  
 <211> 508  
 <212> DNA  
 <213> Homo sapien

<400> 11

gcctgcgagt	ctgtcctggt	ctagcagtga	agcctggttt	gccatgcacg	gggctgggaa	60
gcaagcagtc	tgggtctgga	gggtggcccc	aaaaaggcca	actctgcagc	tccacagcca	120
catgggggag	gctgccacag	gtcacactcc	ttagccggat	cccctcctga	agaaaagcat	180
ctgagctgac	cagcgagact	gcagtggggc	agacactatt	aaatttgctg	gcacagctct	240
gagtaccccc	tcgcttcac	cctgacttca	tccegcagta	gctctcagcc	ctcccagccc	300
ctgcagggcc	acgtctttct	ctattgccgg	tcagcgtgtg	tgtgcacaaa	gcccctaagg	360
tttcatgtgt	acacaccggg	gctaagtgtt	ttttacaccc	ttgtgcatct	ctcggcctgg	420

ggctcctgtg caggttgccc tgagagttgg gtttttagtt caaaaagaag gaacacagat 480  
gactactctg ctggcgacac ggccacta 508

<210> 12  
<211> 836  
<212> DNA  
<213> Homo sapien

<400> 12  
cagcgggact cagctggggt gatttcgccc cccatctccg ggggaatgtc tgaagacaat 60  
tttggttacc tcaatgaggg agtggaggag gatacagtgc tactaccaac tagtggataa 120  
aggccaggga tgctgctcaa cctcctacca tgtacaggac gtctcccat tacaactacc 180  
caatccgaag tgtcaactgt gtcaggacta agaaaccctg gttttgagta gaaaagggcc 240  
tggaagagag ggagccaaca aatctgtctg ctccctcaca ttagtcattg gcaaataagc 300  
attctgtctc tttggtgct gcctcagcac agagagccag aactctatcg ggcaccagga 360  
taacatctct cagtgaacag agttgacaag gcctatggga aatgcctgat gggattatct 420  
tcagcttgtt gagcttctaa gtttctttcc ctctattcta ccctgcaagc caagttctgt 480  
aagagaaatg cctgagttct agctcagggt ttcttactct gaatttagat ctccagacc 540  
ttcctggcca caattcaaata taaggcaaca aacatatacc ttccatgaag cacacacaga 600  
cttttgaaag caaggacaat gactgcttga attgaggcct tgagggaatga agctttgaag 660  
gaaaagaata ctttgtttcc agcccccttc ccacactytt catgtgttaa ccaactgcct 720  
cctggacctt ggagccacgg tgactgtatt acatgttggt atagaaaact gattttagag 780  
ttatgatcgt tcaagagaat gattaaatat acatttctta cmccaaaaaa aaaaaa 836

<210> 13  
<211> 530  
<212> DNA  
<213> Homo sapien

<400> 13  
gctctatttg catatttatt tcagatgttt cgagtaaact tgaaaagggt aggcacgaag 60  
caatttgttg ctgcttgtea cccccaagtc cccgtggagg ttctgtattt taagaaacag 120  
tgcggtgagt gtacagattt tatttatgct taatttaatg gggctgtgta atactgggtg 180  
acttcttaag acttttttga gacatgggat ccaattttta tattaacttt taatgggtgat 240  
ggggtaatct ataacacatc ataagggttt attcatatat atacagggtt ttaagaatta 300  
agaggatgct gggctctggt cttggcttgg aagattctat ttaattgaaa ctctctgttc 360  
agaaagcaat aactttgtct cgttcctgtt gggctgaacc ctaagggtgag tgtgcagtac 420  
agtgtgtgtg ggtgaaatgg agatttggaa ttgaactctc tgctgtgtaa tgttccccaa 480  
ataattgttg tgtgtatgat acgtgtataa taaaagtatt cttgttagaa 530

<210> 14  
<211> 809  
<212> DNA  
<213> Homo sapien

<220>  
<221> misc\_feature  
<222> (1)...(809)  
<223> n = A,T,C or G

<400> 14  
ataattttta aaaaaaagta aaagagaaaa aaaaaaagac agaattcttcg ctcatgcact 60  
tacctgcttt ttcttctgtt tactcaggaa taactgtgtt tatcaatatt tgggctcttc 120  
accacaaccc ctatttctgg gaagaccctc aggtctttta ccccttgaga ttctccaggg 180  
aaaattctga aaaaatacat ccctatgcct tcataccatt ctgagctgga ttaaggaact 240

gcattgggca	gcattttgcc	ataattgagt	gtaaagtggc	agtggcatta	actctgctcc	300
gcttcaagct	ggctccagac	cactcaaggc	ctccccagcc	tggtcgtcaa	gttgteectca	360
agtccaagaa	tggaatccat	gtgtttgcaa	aaaaagtttg	ctaattttaa	gtcctttcgt	420
ataagaatta	atgagacaat	tttcctacca	aaggaagaac	aaaaggataa	atataataca	480
aaatatatgt	atatggttgt	ttgacaaatt	atataactta	ggatacttct	gactgggtttt	540
gacatccatt	aacagtaatt	ttaattttct	tgctgtatct	ggtgaaaccc	acaaaaacac	600
ctgaaaaaac	tcaagctgac	ttccactgcy	aagggaaatt	attggtttgt	gtaactagtgy	660
gtagagtggc	tttcaagcat	agtttgatca	aaactccact	cagtatctgc	attantttta	720
tctctgcaaa	tatctgcatg	atagctttat	tctcagttat	ctttccccat	aataaaaaat	780
atctgccgca	aaaaaaaaaa	aaaaaaaaaa				809

&lt;210&gt; 15

&lt;211&gt; 866

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 15

gtcacgctca	gtgcatccca	ggcagctctt	agtgtggagc	agtgaactgt	gtgtgggttcc	60
ttctacttgg	ggatcatgca	gagagcttcg	cgtctgaaga	gagagctgca	catgttagcc	120
acagagccac	ccccaggcat	cacatgttgg	caagataaag	accaaataga	tgacctgcga	180
gctcaaatat	taggtggagc	caacacacct	tatgagaaag	gtgtttttta	gctagaagtt	240
atcattcctg	agaggtaccc	atttgaacct	cctcagatcc	gatttctcac	tccaatttat	300
catccaaaca	ttgattctgc	tggaaggatt	tgtctggatg	ttctcaaatt	gccacaaaaa	360
gggtgcttga	gaccatccct	caacatcgca	actgtgttga	cctctattca	gctgctcatg	420
tcagaaacca	accctgatga	cccgtctcatg	gctgacatat	cctcagaatt	taaatataat	480
aagccagcct	tctcaagaa	tgccagacag	tggaacagaga	agcatgcaag	acagaaacaa	540
aaggctgatg	aggaagagat	gcttgataat	ctaccagagg	ctggtgactc	cagagtacac	600
aactcaacac	agaaaaggaa	ggccagtcag	ctagtaggca	tagaaaagaa	atttcacctc	660
gatgttttagg	ggacttgtcc	tggttcatct	tagttaatgt	gttcttttgc	aagggtgatyt	720
aagttgccta	ccttgaattt	ttttttaaat	atatttgatg	acataatttt	tgtgtagttt	780
atttatcttg	tacatatgta	ttttgaaatc	ttttaaacct	gaaaaataaa	tagtcattta	840
atgttgaaaa	aaaaaaaaaa	aaaaaa				866

&lt;210&gt; 16

&lt;211&gt; 500

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 16

atctcagaag	ctttttaaca	gttggtgaaa	tgtgcttgct	caacaaagca	tcctaacagg	60
gtcgttccca	tacacatttg	acctggtcag	ccttttccag	gtgaatagcc	ccagttctga	120
cataaagaaa	gttttatttg	tattttacta	ctgtttggct	aattttgata	tataactggg	180
tacaaacaga	gccttactat	ttattagtgg	ggaaatgatt	ttaagaccgt	ccttttcagt	240
atttaattct	gacagatctg	catccctgtt	ttgttttggg	ttatttctgt	tttggaataa	300
gctgtctcat	ttaaaactgt	tggaatatagc	tggaatcctgg	ataggaaaat	gaaattattt	360
tttcattgtg	ttttttaatt	ggggtgatcc	aaagctggca	ccttcaggca	cattgggtctc	420
atagccatta	ctgtttttat	tgccttctta	agatcctgtc	ttcagctggg	tcagagaaaa	480
cttcttgact	aaaactggct					500

&lt;210&gt; 17

&lt;211&gt; 732

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 17



tactgatctc	tgttctagt	agtgggtcaa	aaagcatatg	catccttata	cgtaactca	60
tcagtccttc	ctcaaggcaa	cctgaggcca	gacaccaaga	aaccaagcgt	atctgtctta	120
aaatgacttg	ttcctgggga	atgccttcaa	ccaaaacaca	gctagtattt	ctatgcccc	180
aatccaatcc	cagtctttca	tgatccatgc	cggcggttgg	gtggggaggg	gaatcattgg	240
ttgggggaag	ggaggaaacc	ccacctccag	cccccgccac	cgggctccct	gggcacccag	300
caagatctgg	ggctgcagag	aacagaagag	ctggtgcaact	taatccagct	ctgcccttgg	360
ggggaggagg	acctgtgtgt	caggctctgc	catgggaacg	agtgtaaacc	gtggctgtct	420
cctgcagtga	gccaccgcgg	caggcacgtt	gactgtttta	ctgacatcac	tcaaaagcta	480
aagcaataac	attctcctgc	gttgcctgag	cagctgttca	tttgtccgcc	agctcctgga	540
ctggatgtgt	gaaaggcatc	acatttccat	tttcctccgt	gtaaatgttt	tatgtgttcg	600
cctactgata	ccattcgttg	cttctattgt	aaatatttgt	catttgtatt	tattatctct	660
gtgttttccc	cctaaggcat	aaaatggttt	actgtgttca	tttgaacca	aaaaaaaaa	720
aaaaaaaaa	aa					732

<210> 18  
 <211> 103  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1) ... (103)  
 <223> n = A,T,C or G

<400> 18	
cnagggggccc	ggggcgggggt ccgcccggccc tgcggggccgc cggtgaaata ccactactct
gatcggtttt	tcactgaccc ggtaaggcgg gggggcgagc ccc
	60
	103

<210> 19  
 <211> 586  
 <212> DNA  
 <213> Homo sapien

<400> 19	
acgcctgcag	gtaccgggtcc ggaattcccg ggtcgactac tactactaaa ttcgcggccg
cgtcgactcc	cagcccttat cgagaaaacc atctccaaag ccaaaggcca gccccgagaa
ccacaggtgt	acaccctgcc cccatcccgg gaggagatga ccaagaacca ggtcagcctg
acctgcctgg	tcaaaggctt ctatcccagc gacatcgccg tggagtggga gagcaatggg
cagccggaga	acaactacaa gaccacgcct cccgtgctgg actccgacgg ctccctcttc
ctctatagca	agctcaccgt ggacaagagc aggtggcagc aggggaacgt cttctcatgc
tccgtgatgc	atgaggtctt gcacaaccac tacacgcaga agagcctctc cctgtccccg
ggtaaatgag	tgcgacggcc ggcaagcccc cgtccccgg gctctcgccg tcgcacgagg
atgcttgga	cgtaccccg ctacatactt ccaggcacc cagcatggaa ataaagcacc
caccactgcc	ctggaaaaaa gtcgacgcgg ccgcgaattt agtagt
	586

<210> 20  
 <211> 646  
 <212> DNA  
 <213> Homo sapien

<400> 20	
ccgggtcgac	tactactact aaattcgccg ccgcgtcgac caacagggga gagtgttaga
gggagaagt	ccccctgttg aagctctttg tgatggcgga gctcagggcc tgatgggtga
cttcgcgggc	gtagactttg tgtttctcgt agtctgcttt gctcagcgtc aggggtgtgc
tgaggctgta	ggtgctgtcc ttgctgtcct gctctgtgac actctcctgg gagttaccg
	240

attggagggc	gttatccacc	ttccactgta	ctttggcctc	tctgggatag	aagttattca	300
gcaggcacac	aacagaggca	gttccagatt	tcaactgctc	atcagatggc	gggaagatga	360
agacagatgg	tgcagccaca	gttcgtttga	tctccagctt	ggccccctgg	ccaaaagtgt	420
acggaggtga	gctgtcgacg	cggccgcgaa	tttagtagta	gtaggcggcc	gctctagagg	480
atccaagctt	acgtacgcgt	gcatgcgacg	tcatagctct	tctatagtgt	cacctaaatt	540
caattcactg	gccgtcgttt	tacaacgtcg	tgactgggaa	aaccctggcg	ttacccaact	600
taatcgctt	gcagcacatc	cccccttcgc	cagctggcgt	aatagc		646

&lt;210&gt; 21

&lt;211&gt; 625

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 21

tttttttttt	tttaccaaaa	ttgaccttta	tttgtgttac	taaagcctgt	tttaacttttg	60
atacaaagta	acatttttagt	acagaaaatc	ccagtctgtc	agctcagtac	ctgtctgtgc	120
acactgtacc	atctcagtc	cactctgcct	gtaacttaaa	aaacagcccc	tacccccaga	180
gggtctgcga	gttaatacct	tgagaatagt	ctacagtttt	tcatagtttg	tctgagctag	240
aaaacttgta	cctgtaaaaac	aaaggacagc	attgaggact	gaaacttgtc	tytttttttg	300
aacaactgtg	caagaaaaata	tatccctttt	taaaaaacat	cagttatggc	taaactacaa	360
tctagtgtct	agaattacaa	agaataaaat	gaaatcaaag	atttctcgct	agtaaaatga	420
aatgtttaga	acagtattaa	aatataggtc	ctaccccaac	gacacttaca	cagagcccag	480
tacagtacct	attattaaca	ggacgcatag	cttaaggagg	aaccacatca	aatcttcagc	540
cagacatac	tagcctcaga	agtgcacaaa	aaaaaaaaag	cccccaaacg	aagacacccc	600
actgagtagg	gtgcatgccg	tgagt				625

&lt;210&gt; 22

&lt;211&gt; 576

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 22

ctctcagttc	ctgcaagcag	ctgagcgcta	tggcattaac	accactgaca	tcttccaaac	60
tgtggacctc	tgggaaggaa	agaatatggc	ctgtgtgcag	cggacgctga	tgaatctggg	120
tgggctggca	gtagcccag	atgatgggct	cttctctggg	gateccaaact	ggttccctaa	180
gaaatccaag	gagaatcctc	ggaactctct	ggataaccag	ctgcaagagg	gcaagaacgt	240
gatcgggtta	cagatgggca	ccaaccgcgg	ggcgtctcag	gcaggcatga	ctggctacgg	300
gatgccacgc	cagatcctct	gatccccccc	caggccttgc	ccctgccctc	ccacgaatgg	360
ttaatatata	tgtagatata	tatttttagca	gtgacattcc	cagagagccc	cagagctctc	420
aagctccttt	ctgtcagggt	gggggggttca	gcctgtcctg	tcacctctga	ggtgcctgct	480
ggcatcctct	cccccatgct	tactaataca	ttcccttccc	catagccatc	aaaactggac	540
caactggcct	cttcctttcc	cctgggacca	aaattt			576

&lt;210&gt; 23

&lt;211&gt; 416

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 23

ggcacgaggg	cggccccgcg	ggtctgtctc	ttgcttcaac	agtgtttgga	cggaaacagat	60
ccggggactc	tcttccagcc	tccgaccgcc	ctccgatttc	ctctccgctt	gcaacctccg	120
ggaccatctt	ctcggccatc	tctgtcttct	gggacctgcc	agcaccgttt	ttgtggttag	180
ctccttcttg	ccaaccaacc	atgagctccc	agattcgtca	gaattattcc	accgacgtgg	240
aggcagccgt	caacagcctg	gtcaatttgt	acctgcaggc	ctcctacacc	tacctctctc	300
tgggcttcta	tttcggaccg	cgatgatgtg	gctctggaag	gcgtgagcca	cttcttccgc	360

gaactggccg aggagaagcg cgagggctac gagcgtctcc tgaagatgca aaacca

416

<210> 24  
 <211> 1037  
 <212> DNA  
 <213> Homo sapien

<400> 24  
 gctttaccta tgggtggtgt tatggcaaca agaacaactt tgaggaagag cwttagtgcc 60  
 tcgagtcttg tcgcggcatc tccattaagg atgtgtttgg cctgaggcgg gaaatcccca 120  
 tccccagcac aggctctgtg gagatggctg tcgcagtgtt cctgggtcatc tgcattgtgg 180  
 tgggtggtagc catcttgggt tactgcttct tcaagaacca gagaaaggac tccacggac 240  
 accaccacca cccaccacc acccctgcc gctccactgt ctccactacc gaggacacgg 300  
 agcacctggt ctataaccac accaccggc cctctgagc ctgggtctca cgggctctca 360  
 cctggccctg ctccctgctt gccaaaggcagg aggcctgggc tgggaaaaac tttggaacca 420  
 gactcttgcc tgtttccag gccactgtg cctcagagac cagggtcca gcccctcttg 480  
 gagaagtctc agctaagctc acgtcctgag aaagctcaaa ggtttggaag gagcagaaaa 540  
 cccttggggc agaagtacca gactagatgg acctgcctgc ataggagtgt ggaggaagt 600  
 ggagttttgt ttcctctgtt caaagctgcc tgtccctacc ccatggtgct aggaagagga 660  
 gtgggggtgt gtcagacctt ggaggcccca accctgtcct cccgagctcc tyttccatgc 720  
 tgtgcgcccc gggctgggag gaaggacttc cctgtgtagt ttgtgctgta aagagtgtgt 780  
 ttttgtttat ttaatgctgt ggcatgggtg aagaggaggg gaagaggcct gtttggccty 840  
 tytgtcccty ytccctcttc ccccaagatt gagctytytg cccttgatca gcccmmccct 900  
 ggccatagacc agcagacaga gccaggagag gtcagctgc attccgcagc cccaccccc 960  
 aagggtttcc aacatcacag cccagccac ccactgggta ataaaagtgg tttgggaaaa 1020  
 aaaaaaaaa aaaaaaa 1037

<210> 25  
 <211> 1144  
 <212> DNA  
 <213> Homo sapien

<400> 25  
 cggcacgagg ctagatgggt ccatgacaat ggtgtgaact acaagattgg agagaagtgg 60  
 gaccgtcagg gagaaaatgg ccakatgatg agctgcacat gtcttgggaa cggaaaagga 120  
 gaattcaagt gtgacctca tgaggcaacg tgttacgatg atgggaagac ataccacgta 180  
 ggagaacagt ggcaagaagga atatctcggg gccatttgcct cctgcacatg ctttggaggc 240  
 cagcgggggt ggcgctgtga caactgccgc agacctgggg gtgaaccagg tcccgaaggc 300  
 actactggcc agtcctacaa ccagtattct cagagatacc atcagagAAC aaacactaat 360  
 gtttaattgcc caattgagtg cttcatgcct ttagatgtac aggctgacag agaagattcc 420  
 cgagagtaaa tcatctttcc aatccagagg aacaagcatg tctctctgcc aagatccatc 480  
 taaactggag tgatgttagc agaccagct tagagttctt ctttctttct taagcccttt 540  
 gctctggagg aagttctcca gcttcagctc aactcacagc ttctccaagc atcaccctgg 600  
 gaggtttctg agggttttct cataaatgag ggctgcacat tgcctgttct gcttcgaagt 660  
 attcaatacc gctcagtatt ttaaatgaag tgattctaag atttggtttg ggatcaatag 720  
 gaaagcatat gcagccaacc aagatgcaaa tgttttgaaa tgatatgacc aaaattttaa 780  
 gtaggaaagt caccctaaaca cttctgcttt cacttaagt tytgggccgc aatactgtag 840  
 gaacaagcat gatcttgta ctgtgatatt ttaaatactc tctcttacct gttatttatc aatttttccc 900  
 atgatcctag taattgccta gaaatatctt tctcttacct gttatttatc aatttttccc 960  
 agtattttta tacggaaaaa attgtattga aaacmcttag tatgcagttg ataagaggaa 1020  
 tttggtataa ttatggtggg tgaaaatttt ttatactgta tgtgccaaag ctttaaaact 1080  
 gtggaaagac aactgtttta ataaaagatt tacattcccc aacttgaaaa aaaaaaaaaa 1140  
 aaaa 1144

<210> 26

<211> 488  
 <212> DNA  
 <213> Homo sapien

<400> 26  
 gcatttttgt ctgtgtctccc tgatcttcag gtcaccacca tgaagttctt agcagtctctg 60  
 gtactcttgg gagtttccat ctttytggtc tctgccaga atccgacaac agctgtctcca 120  
 gctgacacgt atccagctac tgggtcctgct gatgatgaag cccctgatgc tgaaaccact 180  
 gctgctgcaa ccaactgagac cactgctgct cctaccactg caaccaccgc tgcttctacc 240  
 actgctcgta aagacattcc agttttaccc aaatgggttg gggatctccc gaatggtaga 300  
 gtgtgtccct gagatggaat cagcttgagt cttctgcaat tggtcacaac tattcatgct 360  
 tcctgtgatt tcatccaact acttaccttg cctacgatat cccctttatc tctaatacagt 420  
 ttatttttyt tcaataaaaa aataactatg agcaacataa aaaaaaaaaa aaaaaaaaaa 480  
 aaaaaaaaaa 488

<210> 27  
 <211> 764  
 <212> DNA  
 <213> Homo sapien

<400> 27  
 agccccaggt gagattgcag ctggtgattt attgagaaac ttccaggttg agtaggcttc 60  
 attcagggca tgtctctccc ccaggcccca tcttcataca cttccgctaa gatgccca 120  
 attcatcccc ctcatgtgc agcttgaaaa accctgacat tgttgctccc tgccgggcat 180  
 acctgccatt gccactgccc atctctgtga tgaggaattc aacaccagat ctctggcagc 240  
 tcttgaggga ttggccagtg gccattctca ggggtgctgt ctagttttcc cagagccccg 300  
 ccaaggtcct ggccaccgtg gcagagcggg gaggttcccc cactctaaag cctactcctt 360  
 ctctctctag cctcttcagg aggtggtggt agcagcagca attccacat caatgtagaa 420  
 gagtcagtgg atggacaggt ggtttcttcc cacaagagag aaatytaagt gtctattgca 480  
 ggagaaacgt cccttgccac tccccactct catcaggcca agtggaggac tggccagagg 540  
 gcctgcacat gcaaactcca gtccctgcct tcagagagct gaaaagggtc cctcggtcctt 600  
 ttatttcagg gctttgcatg cgtcttatcc cccctctgcc tctccccmcc ttytttgag 660  
 caaggagatg cagctgtatt gtgtaacaag ctcatattgta cagtgtctgt tcatgtaata 720  
 aagaattact tttccttttg caaaaaaaaaa aaaaaaaaaa aaaa 764

<210> 28  
 <211> 802  
 <212> DNA  
 <213> Homo sapien

<400> 28  
 accatcactt gccgggcaag tcagagtatt aacaactatt taaattggta tctgcagaaa 60  
 ccagggaag cccctaacct cctgatctat gctgcatcca ctttgcaaag tgggggtcca 120  
 tcaagggtca stggcagtga atctgggaca gatttctactc tcaccatcag cagtctgcaa 180  
 cctgaagatt ttgcaactta ctactgycaa cagagttaca ctacccatt cactttcrgm 240  
 cctgggacca aagtggatgt caaacgaact gtggctgcac catctgtctt catcttcccg 300  
 ccctctgatg agcagwtgaa atctggaact gcctctgttg tgtgctgct gaataacttc 360  
 tatcccagag aggccaaagt acagtggaag gtggataacg ccctccaatc gggttaactcc 420  
 caggagagtg tcacagagca ggacagcaag gacagcacct acagcctcag cagcaccctg 480  
 acgtgagca aagcagacta cgagaaacac aaagtctacg cctgsaaagt caccatcag 540  
 ggcctgagct cgcccgctac aaagagcttc aacaggggag agtgtagag ggagaagtgc 600  
 cccacactgc tcctcagttc cagcctgacc ccctccatc ctttggccty tgaccctttt 660  
 tccacagggg acctaccctt attgcggtcc tccagctcat ctttcacctc accccctcc 720  
 tcctccttgg ctttaattat gctaagtgtg gaggagaatg aataaataaa gtgaatcttt 780  
 gcaaaaaaaaaa aaaaaaaaaa aa 802

<210> 29  
 <211> 620  
 <212> DNA  
 <213> Homo sapien

<400> 29  
 ctcgagccag aattcggtcg aggagaactg gtataggagc tccaaggaca agaaacacgt 60  
 ctggctagga gaaactatca atgctggcag ccagtttgaa tataatgtag aaggagtgc 120  
 ttccaaggaa atggctaccc aacttgcctt catgcgctg ctggccaact atgcctctca 180  
 gaacatcacc taccactgca agaacagcat tgcatacatg gatgaggaga ctggcaacct 240  
 gaaaaaggct gtcattctac agggctctaa tgatgttgaa cttgttgctg agggcaacag 300  
 cagggttact tacactgttc ttgtagatgg ctgctctaaa aagacaaatg aatggggaaa 360  
 gacaatcatt gaatacaaaa caaataagcc atcacgcctg cccttccttg atattgcacc 420  
 tttggacatc ggtgggtgctg accaggaatt ctttgtggac attggcccag tctgtttcaa 480  
 ataatgaac tcaatctaaa ttaaaaaaga aagaaatttg aaaaaacttt ctctttgcc 540  
 tttcttcttc ttctttttta actgaaagct gaatccttcc atttcttctg cacatctact 600  
 tgcttaaat gtgggcaaaa 620

<210> 30  
 <211> 644  
 <212> DNA  
 <213> Homo sapien

<400> 30  
 aattagagtt aaattcacia tattttttaa gaaaacatta tgggaagatg attcatttca 60  
 aaccacmcg cmatthaaca taaaacactt gkcaagctga gtagactgkt ttcttatgk 120  
 aaccacaaaa tattttctct gaaatctaca cttagttaa aaacagagat gggattttgc 180  
 atattagctt gaaaataagt atatgatgat gatattaggk gcccactagc acctagtttt 240  
 tacagctttg cattgkcacc ccactactgc cagggaccca gcccaggca tacacagatg 300  
 aaaggacagt ttcaccttct tggcaaaaac cttcagaaca attgtcaaca tactctcaa 360  
 tgtctttccc actcagaaat gaggagcaag gtgtatgact ttagattcaa gaagtatatg 420  
 gggctaaata tctttaaag tataactctg gacaatgtac ttagggacct actacttact 480  
 caaataggg tagtagcatt aataactaaa tagggcgtga gtcatgagaa agattccctg 540  
 catctcttgc caagtattaa cactagtttt actaataagc aacggggtag ataactttga 600  
 agattgcac aacttcaaca caaaaaatg tgaatatccc tcgt 644

<210> 31  
 <211> 674  
 <212> DNA  
 <213> Homo sapien

<400> 31  
 cttctggaat atgctagggc agaagtatgt caagtatgtc acatgaagaa aatgtgaaat 60  
 tcaagagtaa tccacacgtg agaaactaga caatgtacat tcatgtgttc tcttgaaagg 120  
 aaagggagag ctgtaagctt cactctgtcc tacaccggag aaaagcagga ataactttac 180  
 cgtggaaata atgtttagct tttatcagag aaaaattgtcc ttctagagca tagagtccca 240  
 aaactcaatt ctggttttcc cctgtttttt tttttttttt ttttcccaac atatgaactg 300  
 cagcatatca ctttttcttt ttgtgcctca ggttcctcag ctgtaaaatt gaaaaatata 360  
 tgtattaata atattattaa taataataat ggtaatgtag tacttgtttg taaagcactt 420  
 tgagatcctt ggttgaaaagg caccatagga gtgccaaagta ttattatgtg gccaaagggg 480  
 ttatttaaac tgtcagttcc caaaggccag gaaagggttg ggtcattttt cttaaagacg 540  
 agctgtaaat atcaactagg cagccaatag tgttgactat gaagatgcaa aactattact 600  
 aggmtgataa aatcatagtt tcttaatggc taccaataag gcaaatatca caataataaa 660  
 caccaaattc cgta 674

<210> 32  
 <211> 713  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(713)  
 <223> n = A,T,C or G

<400> 32  
 gccagaattc ggtcaggac agctataaaa gaaagctgtt atttactgta gttgtagatg 60  
 tattcactac akaatcctac atttttcagc aggccacagt ccancctaat cctataatat 120  
 ccttgaaaag aaactattaa aaaggataga catttctgat gtaagtaaaa cccccaagca 180  
 acttaatatc catctgtcag tcatcaactt tccccctaga tttttttttt aactagtctt 240  
 gaaagtgtca agaataagctt cccttttcac cctacctagg tccctcctt gctggcataa 300  
 aggccaaagca gtgaaatgaa ccttgggtta gagttatggt aggcaaaaag aaaaggagaa 360  
 tttaggaaat aaatatgcta taaacaaggt acagtacagt gagagttaac aatcctagga 420  
 atcctgcctg tcacaagctc ccagttccct gtcattttat ctcatgatt agaactgac 480  
 ttccatttaa ctattacaga aagcagtaac tattgcaact atctaatagt tcacacaaaa 540  
 gaattaaaag cttaaaaata atttttagga aacacaatat tcaaaatcta aacacactga 600  
 taaattatta agattaagat tatwtatgtg ataatgaaa tctcctacca atccatccag 660  
 cctttaccag ggaagaaaag caattatttc atttcagata gaaatacaaa aaa 713

<210> 33  
 <211> 698  
 <212> DNA  
 <213> Homo sapien

<400> 33  
 tcgagggtc caccgaaaga gtcactggaa ctggaggacc cgtcttctgg gctgggtgtg 60  
 accaagcagg atctgggccc agtccccatg tgagagcagc agaggcggtc ttcaacatcc 120  
 tgccagcccc acacagctac agctttcttg ctcccttcag cccccagccc ctcccccatc 180  
 tcccacctg tacctcatcc catgagaccc tgggtgcctgg ctctttctgc acccttggac 240  
 aagacaaacc aagtcggaac agcagataac aatgcagcaa ggccctgctg cccaatctcc 300  
 atctgtcaac aggggcgtga ggtcccagga agtggccaaa agctagacag atccccgttc 360  
 ctgacatcac agcagcctcc aacacaaggc tccaagacct aggttcatgg acgagatggg 420  
 aaggcacagg gagaaggat aaccctacac ccagaccca ggctggacat gctgactgtc 480  
 ctctccccct cagcctttgg ccttggcttt tctagcctat ttacctgcag gctgagccac 540  
 tctctccctt ttccccagca tcaactccca aggaagagcc aatgttttcc acccataatc 600  
 cttctgtccg acccctagtt cctctgtctc agccaagcta gwtatcagct ttcaggggca 660  
 tggttcacat tagaataaaa ggtagtaatt agaaaaaa 698

<210> 34  
 <211> 605  
 <212> DNA  
 <213> Homo sapien

<400> 34  
 aaggtgccat cattcttgag gaggaagtag cgtggccgcc aggtcttgat gtactcccc 60  
 cgtttgtgca gccaccctc cttcacaata gccacgtcgc tcatgggtgc cgaggctccc 120  
 gcgacgtca cgcgtcctc tcaggctggc gctccccag cccagctggc ctggccacag 180  
 cctctgatgc accagctgac aggctgcctc ctccaggcag cccctttgac ttctttgacc 240  
 caggctggct cggccttccc taagccctg gtgacagatg gccccgtttg ctctccctgt 300

```

ccatggtggt cctacccatg gatcctggga cagggcacag ggctcctccc tgctccccag 360
actaggaag caaagaaatt caaacatgag gaagacagga ccaggatgca ggccactggc 420
gcaaacggga gtccagagcc ctccagcgca agccccaaaa cctcctggga gaaaccccag 480
gccccctcta aaccacagcg cccctgccgg tctgaatctg gttcattcat ttggccaaca 540
tgtaccgggc gtctcaggtt ttgccaggcc cagtgtctgg cgctggagac aaagagcctc 600
gtrcc 605

```

```

<210> 35
<211> 753
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(753)
<223> n = A,T,C or G

```

```

<400> 35
atggagacgg ctctgcagca gtacactcta gaaccatcag aaaaaccttt tgttctcaag 60
tctgtgcccc tggccacggc gcccatggca gagcagagaa cagaaagtac ccccatcaca 120
scagtcaaac agcctgagaa agtggcagct accaggcagg agatcttcca ggagcagttg 180
gcagcagtc cagagttccg cggctcttggg cccctcttca agtcctcgcc tgagcccgtg 240
gccctcaccg agtcagagac ggagtatgtc atccgctgca ccaaacacac ctccaccaac 300
cacatggttt ttcagtttga ctgcacaaac acactcaatg accagacctt ggagaatgtc 360
acagtgcaga tggagcccac tgaggcctat gaggtgctct gttacgtgcc tgcccggagc 420
ctgccctaca accagcccgg gacctgtctac acactggtgg cactgcccac agaagacccc 480
acagctgtgg cctgcacatt cagctgcatg atgaagttca ctgtcaagga ctgtgatccc 540
accactgggg agactgatga cgaaggctat gaggatgagt atgtgctgga agatctggaa 600
gttactgtag ctgatcacat tcaaaaggtc atgaaactga acttcgaagc agcctgggat 660
gangtagggg atgaatttga gaaggaggaa acgttcacct tgtctaccat caagacactt 720
gaagaggctg ggggnaatat tgggaagtct tgg 753

```

```

<210> 36
<211> 433
<212> DNA
<213> Homo sapien

```

```

<400> 36
ggggcttgca atcatttgaa ttgttctgtt tcacaataaa ggagattcac tgggttctgc 60
attttcagga ttcaatagaa ctgcttcatt aaaaaataa tccttagcaa gcattcgaat 120
cctaactgct ttgatgcact tgccctcggg cacctgtcat ttccaatatg gtaggtgtca 180
aagtcaaaag tatttactgg gagaaaaaag agaggagtgg ttgtagaagt ctccctaaat 240
cagacatgtc aagcaatcag ccaacgtggt gtatttctca ttcaatattt tagtgtgaat 300
tgagacactg agataaagac atcgtgcaga gataaatggg gatacagtta aatgtagcaa 360
ctcttgagtt cattttttcc cactgtagca aaattaatgc tttctcttta ttgaaataaa 420
ttgctcattc ctc 433

```

```

<210> 37
<211> 601
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (1)...(601)

```

<223> n = A,T,C or G

<400> 37

aggagtcccc	cggtcgcccc	tcacctgtg	gcaagtacat	ctcctgcgcc	gagtgcctga	60
agttcgaaaa	gggccccttt	gggaagaact	gcagcgcggc	gtgtccgggc	ctgcagctgt	120
cgaacaaccc	cgtgaagggc	aggacctgca	aggagaggga	ctcagagggc	tgctgggtgg	180
cctacacgct	ggagcagcag	gacgggatgg	accgctacct	catctatgtg	gatgagagcc	240
gagagtgtgt	ggcaggcccc	aacatcgccg	ccatcgctcg	gggcaccgtg	gcaggcatcg	300
tgctgatcgg	cattctcctg	ctggtcatct	ggaaggctct	gatccacctg	agcgacctcc	360
gggagtacag	gcgctttgag	aaggagaagc	tcaagtccca	gtggaacaat	gataatcccc	420
ttttcaagag	cgccaccacg	acggatcatga	accccaagtt	tgctgagagt	taggagcaact	480
tggtgaaaaa	aanggccgtc	aggaccaccc	atgtctgccc	atacgcgggc	ganacatggc	540
ttggcacact	nttgangatg	taccatttac	caaatcnan	tttttcccct	tnaatnanct	600
g						601

<210> 38

<211> 713

<212> DNA

<213> Homo sapien

<220>

<221> misc\_feature

<222> (1)...(713)

<223> n = A,T,C or G

<400> 38

gcttccgcct	attgattctg	ctcatttata	caataatgag	gagcagggtg	gactggccat	60
ccgaagcaag	attgcagatg	gcaktgtgaa	gagagaagac	atattctaca	cttcaaagct	120
ttggtgcaat	tcccatcgac	cagagtgtgt	cgcaccagcc	ttggaaaggt	cactgaaaaa	180
tcttcaattg	gattatgttg	acctctacct	tattcathtt	ccagtgtctg	taaagccagg	240
tgaggaagtg	atcccaaaag	atgaaaatgg	aaaaatacta	tttgacacag	tggatctctg	300
tgccacgtgg	gaggccgtgg	agaagtgtaa	agatgcagga	ttggccaagt	ccatcggggt	360
gtccaaactt	aaccgcaggc	agctggagat	gatcctcaac	aagccagggc	tcaagtacaa	420
gcctgtctgc	aaccaggtgg	aatgtcatcc	ttacttcaac	cagagaaaaa	tgctggattt	480
ctgcaagtca	aaagacattg	ttctgggtgc	ctatagtgtc	ctgggatccc	accgagaaga	540
accatgggtg	gacccgaact	ccccgggtgc	cttggaggac	ccagtccttt	gtgccttggc	600
aaaaaagcac	aagcgaaccc	cagccctgat	tgccctgcgc	taccactaca	gcgtgggggt	660
gtggtcctgg	ccaagagcta	caatgagcag	cgcatcanac	agaacgtgca	ggg	713

<210> 39

<211> 451

<212> DNA

<213> Homo sapien

<400> 39

gaaagaactg	actgaaacgt	ttgagatgaa	gaaagttctc	ctcctgatca	cagccatctt	60
ggcagtggtc	gttggtttcc	cagtctctca	agaccaggaa	cgagaaaaaa	gaagtatcag	120
tcgacagcga	tgaattagct	tcagggtttt	ttgtgttccc	ttacccatat	ccatttcgcc	180
cacttccacc	aattccattt	ccaagatttc	catggtttag	acgtaatttt	cctattccaa	240
tacctgaatc	tgcccctaca	actccccttc	ctagcgaaaa	gtaaacaaga	aggaaaagtc	300
acgataaacc	tggtcacctg	aaattgaaat	tgagccactt	ccttgaagaa	tcaaaattcc	360
tgtaataaaa	agaaaaacaa	atgtaattga	aatagcacac	agcattctct	agtcaatatc	420
tttagtgatc	ttctttaata	aacttgaaag	c			451

<210> 40



<211> 778  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(778)  
 <223> n = A,T,C or G

<400> 40  
 accgcacacct agccgccgac tcacacaagg caggtgggtg aggaaatcca gagttgccat 60  
 ggagaaaatt ccagtgtcag cattcttgct ccttggtggc ctctcctaca ctctggccag 120  
 asataccaca gtcaaacctg gagccaaaaa ggacacaaag gactctcgac ccaaactgcc 180  
 ccagaccctc tccagagggt ggggtgacca actcatctgg actcagacat atgaagaagc 240  
 tctatataaa tccaagacaa gcaacaaacc cttgatgatt attcatcact tggatgagtg 300  
 cccacacagt caagctttaa agaaagtgtt tgctgaaaat aaagaaatcc agaaattggc 360  
 agagcagttt gtcctcctca atctgggtta tgaaacaact gacaaacacc tttctcctga 420  
 tggccagtat gtcccagga ttatgtttgt tgacccatct ctgacagtta gagccgatat 480  
 cactggaaga tattcaaate gtctctatgc ttacgaacct gcagatcagc tctgttgctt 540  
 gacaacatga agaaagctct caagttgctg aagactgaat tgtaaagaaa aaaaatctcc 600  
 aagcccttct gtctgtcagg ccttgagact tgaaaccaga anaagtgtga naaagactgg 660  
 ctagtggggg aagcattagt ggaacacact ggattanggt tatgggtttn aatggttacc 720  
 aacaacctat ttttttttaa naaaaaacc aaagttttta aaaaaaattt ngggtttt 778

<210> 41  
 <211> 696  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(696)  
 <223> n = A,T,C or G

<400> 41  
 agtgatgccg gcagatcatc tggtttcagc gggacnacgt atcttgcagt tggctctggca 60  
 tggaccncag ggctgtgggg acttggggta cagtaatcaa gtaatcccct tttccagaat 120  
 gcattaaccc actccctttg acctcacgat ggggcaggtc cccaagtgtg caagctcagt 180  
 attcatgatg gtgggggatg gagtgtcttc cgaggttctt gggggaaaaa atattgtagc 240  
 atatttaagg gaggcaatga accctctccc ccnctcttc cctgcccata tctgtctcct 300  
 agaatcttat gtgctgtgaa taataggcct tcwctgcccc tccagttttt atagacstga 360  
 ggttccagtg tctcctggta actggaacct ytcctgaggg ggaatcctgg tgctcaaatt 420  
 accctccaaa agcaagtagc caaagccgtt gccaaacccc acccataaat caatgggccc 480  
 tttatttatg mcgactttat ttattcctaat atgattttat agtatttata tatattgggt 540  
 cgtctgcttc ccttgatatt ttmttccttt ttttgaata ttgaaaasgg sgatataatt 600  
 attataagta gactataata tatttagtaa tatatattat taccttaaaa gtctattttt 660  
 gtgttttggg cattttttaa taaacartct gagtgt 696

<210> 42  
 <211> 509  
 <212> DNA  
 <213> Homo sapien

<400> 42  
 ctcaggtgga aaaggaggga gctactctca ggctgcgtgc agcgacagtg cccagggctc 60

tgatgtgtct	ctcacagctt	gaaaagcctg	agacagctgt	cttgtgaggg	actgagatgc	120
acgatttctt	cacgcctccc	ctttgtgact	tcaagagcct	ctggcatctc	tttctgcaaa	180
ggcacctgaa	tgtgtctgcg	tccctgttag	cataatgtga	ggaggtggag	agacagccca	240
cccttgtgtc	cactgtgacc	cctgttccca	tgctgacctg	tgtttctctc	ccagtcattc	300
ttcttgttcc	acaagagggtg	gggctggatg	tctccatctc	tgtctcaact	ttacgtgcac	360
tgagctgcaa	cttcttactt	ccctactgaa	aataagaatc	tgaatataaa	tttggtttct	420
caaataattg	ctatgagagg	ttgatggatt	aattaaataa	gtcaattcct	ggaatttggg	480
gagagcaaat	aaagacctga	gaaccttcc				509

&lt;210&gt; 43

&lt;211&gt; 388

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 43

gcaggatttc	ttcactcttc	ccctttgtga	cttcaagagc	ctctggcatc	tctttctgca	60
aaggcacctg	aatgtgtctg	cgccctgtt	agcctaattg	gaggaggtgg	agagacagcc	120
cacccttgtg	tccactgtga	cccctgttcc	catgtgacc	tggttttctc	ccccagtcac	180
ctttcttgtt	ccagagaggt	ggggctggat	gtctccatct	ctgtctcaac	tttatgtgca	240
ctgagctgca	acttcttact	tccctactga	aaataagaat	ctgaatatac	atttggtttc	300
tcaaawattt	gctatgagag	gttgatggat	taattaaata	agtcaattcc	tggaatgtga	360
gagagcaaat	aaagacctga	gaaccttc				388

&lt;210&gt; 44

&lt;211&gt; 565

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 44

ggaaaagagg	gttcctctcc	ttgtgwtgt	ctcttccccc	cacccctaata	tcttctgctc	60
tgtttgggaa	gacgtggagg	aaaaggtagc	ttctgcccc	accgctctta	ccccaccaa	120
ctctcgctg	ttggagagaa	ggggccctcc	cagcacaag	ttgcatctc	ccccctaata	180
ttattctaata	ttattaactt	tgacccaccc	tttctgagcc	tgacgcttc	ccgtgtggcc	240
tgagggtctg	cgagttagct	gccccagccc	ctcccagccc	ttgcccagcc	tgggggagtg	300
gggaaggctt	gggcatggcc	ccgttggagg	ttgatttgc	gttttgttcc	ttgtctttgt	360
gttctgtggt	acttgcctgag	agaaaagaaa	agttagccaa	gcagaaggag	gtgggaaaac	420
ggacccaaac	cccagtgctg	cctgccccga	tgcccttctc	ttagtgggtg	gaaaccctta	480
tctygcaaag	tgaatgtgtc	cccttcccc	cccthtagt	tatttcacag	aaaacaaaac	540
ctcccaataa	aacgggtgaa	gcttg				565

&lt;210&gt; 45

&lt;211&gt; 661

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 45

gttgaggaga	tgggatgtcc	cagatgatag	ggctcctggg	atttcagacc	caataccagc	60
aggactccag	tcacctctac	cccakttctc	caggacacag	cgctcccaac	tctgagtgc	120
gtcccacctc	tggtccttgc	agcacaacca	acgtgggaat	cacacctcc	agacctccca	180
cagctccacc	ccagactggg	cgccggccct	gcctccattt	cagctgtgac	aacctcagag	240
ccgtgttggc	ccaagcatga	caaggacgta	tgaaaacttc	cagtacttgg	agaataagg	300
gaaagtccag	gggttttaaa	atggggccact	tccctctccag	tccctcctgc	agcgtctccg	360
ctctgggccc	tgccatctcc	tgctgtccct	gggcctcggc	ctgctgctgc	tggtcatcat	420
ctgtgtgggt	ggattccaaa	attccaaatt	tcagagggac	ctggtgaccc	tgagaacaga	480
ttttagcaac	ttcacctcaa	acactgtggc	ggagatccag	gcactgactt	cccagggcag	540

```

cagcttgga gaaacgatag catctctgaa agctgaggtg gagggtttca agcaggaacg      600
gcaggcagtt cattctgaaa tgctcctgcg agtcacagcag ctggtgcaag acctgaagaa      660
a                                                                              661

```

```

<210> 46
<211> 453
<212> DNA
<213> Homo sapien

```

```

<400> 46
tgctctctgt gtgtgcctgc tgaataactt ctatcccaga gaggccaaag tacagtggaa      60
gggtggataac gccctccaat cgggtatctc ccaggagagt gtcacagagc aggacagcaa      120
ggacagcacc tacagcctca gcagcaccct gacgctgagc aaagcagact acgagaaaca      180
caaagtctac gcctgcgaag tcacccatca gggcctgagc tcgcccgtca caaagagctt      240
caacagggga gagtgttaga gggagaagtg ccccccactg ctccctcagtt ccagcctgac      300
cccctcccat cctttggcct ctgacccttt tccacaggg gacctacccc tattgcgggc      360
ctccagctca tctttcacct cccccccctc ctccctcctg gctttaatta tgctaattgt      420
ggaggagaat gaataaataa agtgaatctt tgc                                          453

```

```

<210> 47
<211> 687
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (1)...(687)
<223> n = A,T,C or G

```

```

<400> 47
cttcggctaa aatgcagaaa tgcattgctgt cagcgttggt atttcacatt caatggagct      60
gaatgttsag gacctcttcc cattgaagct ataatttatt tggaccaagg aagccctgaa      120
atgaattcaa caattaatat tcatcgact tcttctgtgg aaggactttg tgaaggaatt      180
ggtgctggat tagtggatgt tgctatctgg gttggcactt gttcagatta cccaaaagga      240
gatgcttcta ctggatggaa ttcagtttct cgcatcatta ttgaagaact accaaaataa      300
atgctttaat tttcatttgc tacctctttt tttattatgc cttggaatgg ttcacttaaa      360
tgacatttta aataagttta tgtatacatc tgaatgaaaa gcaaagctaa atatgtttac      420
agaccaaagt gtgatttcac actgttttta aatctagcat tattcatttt gcttcaatca      480
aaagtggttt caatatTTTT tttagtttgt tagaatactt tcttcatagt cacattctct      540
caacctataa tttggaatat tgttgtggct tttgtttttt ctcttagtat agcattttta      600
aaaaaatata aaagctacca atctttgtca atttgtaaat gttaaanaat tttttttata      660
tctggtnaaa taaaaaatta ttttccc                                          687

```

```

<210> 48
<211> 944
<212> DNA
<213> Homo sapien

```

```

<400> 48
ggcacgagga gcatggacat gaggggtccc gctcagctcc tggggctcct gctactctgg      60
ctccgaggtg ccagatgtga catccagatg acccagtctc catcttccct gtctgcttct      120
gtaggagaca gagtcccat cacttgccgg gcaagtcaga acattaacaa cttcttaatt      180
tggtatcagc aaaaaccagg gagacccctc cagctcctga tctatcttac atctaatttg      240
caaagtgggg tcccatctag gttcagtggc agtggatctt ggacagattt cactctcacc      300
atcaccggtc tgcagcctga agattttgca acttactact gtcaacagag tcacagtacc      360

```

ccgtacacct	tccgccaggg	gaccaagctg	gagatcaaac	gaactgtggc	tgcaccatct	420
gtcttcatct	tcccgccatc	tgatgagcag	ttgaaatctg	gaactgcctc	tgttgtgtgc	480
ctgctgaata	acttctatcc	cagagaggcc	aaagtacagt	ggaagggtga	taacgccctc	540
caatcgggta	actcccagga	gagtgtcaca	gagcaggaca	gcaaggacag	cacctacagc	600
ctcagcagca	ccctgacgct	gagcaaagca	gactacgaga	aacacaaaact	ctacgcctgc	660
gaagtcaccc	atcaggggcct	gagctctccc	gtcacaaaga	gcttcaacag	gggagagtgt	720
tagagggaga	agtgccccca	cctgtctctc	agttccagcc	tgacccccctc	ccatcctttg	780
gcctctgacc	ctttttccac	aggggaccta	cccctattgc	ggcctccag	ctcatctttc	840
acctcaccce	cctcctcctc	cttggcttta	attatgctaa	tgttgaggga	gaatgaataa	900
ataaagtga	tctttgcacc	tgtaaaaaaa	aaaaaaaaaa	aaaa		944

&lt;210&gt; 49

&lt;211&gt; 449

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 49

ggcacgaggg	cttcgaggaa	ccagccctga	tacctgtgac	cctctctgta	cgctaccctc	60
ctgaagtgtc	catctccggc	tatgatgaca	actggtagct	cgcccgtagt	gatgccaccc	120
tgagctgtga	cgtccgcagc	aaccagagc	ccacgggcta	tgactggagc	acgacctcag	180
gcaccttccc	gacctccgca	gtggcccagg	gctcccagct	ggcatccac	gcagtggaca	240
gtctgttcaa	taccaccttc	gtctgcacag	tcaccaatgc	cgtgggcatg	ggccgcgttg	300
agcaggtcat	ctttgtccga	gaaaccccca	gggcctcgcc	ccgagatgtg	ggcccgctgg	360
tgtggggggc	cgtggggggg	acactgctgg	tgctgctgct	tctggctggg	gggtccttgg	420
ccttcaccc	gctgaggggtg	aggaggagg				449

&lt;210&gt; 50

&lt;211&gt; 703

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 50

ggcacgaggg	tgagcccatg	ccccacccaa	ccctgccagc	tgagcaaagg	acagtcttac	60
agcgtcaatg	tcaccttcac	cagcaatatt	cagtctaaaa	gcagcaaggc	cgtgggtgcat	120
ggcatcctga	tgggcgtccc	agttcccttt	ctcatctcctg	agcctgatgg	ttgtaagagt	180
ggaattaaact	gccctatcca	aaaagacaag	acctatagct	acctgaataa	actaccagt	240
aaaagcgaat	atccctctat	aaaactgggtg	gtggagtgcc	aaacttcagg	tgacaaaac	300
caaagtctct	tctgctggga	aatcccagta	cagatcggtt	ctcatctcta	agtgcctcat	360
tgagttcgg	gcctctggcc	aatgagctctg	ctgagactct	tgacagcacc	tccagctctg	420
ctgcttcaac	aacagtgact	tgctctccaa	tggtatccag	tgattcgttg	aagaggaggt	480
gctctgtagc	agaaactgag	ctccgggtgg	ctgggttctca	gtggttgtct	catgtctctt	540
tttctgtctt	aggtgggttt	attaaatgca	gcacttggtt	agcagatgtt	taattttttt	600
tttaacaaca	ttaacttggtg	gcctctttct	acacctggaa	atttactctt	gaataaataa	660
aaactcggtt	gtcttgtcaa	aaaaaaaaaa	aaaaaaaaaa	aaa		703

&lt;210&gt; 51

&lt;211&gt; 963

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 51

caatgaggct	ccctgctcag	ctcctggggc	tgctaattgct	ctgggtctct	ggatccagt	60
gggatattgt	gatgactcag	tctccactct	cccagcccg	cacccctgga	gagccggcct	120
ccatctcctg	caagtctagt	cagagcctcc	tacatagtaa	tggatacagt	tatttggatt	180
ggtacctgca	gaagccagga	cagtcctcac	agctcctgat	ctatttggct	tctaagcggg	240

```

cctccggggt ccctgacagg ttcagtggca gtggatcagg cacagatttt acactgagaa 300
tcagtagagt ggaggctgag gatgttggct tttattactg catgcaaact ctacagaccc 360
cgctcacttt cggcggaggg accaaggtgg agatcaaacg aactgtggct gcaccatctg 420
tcttcatctt cccgccatct gatgagcagt tgaaatctgg aactgcctct gttgtgtgcc 480
tgctgaataa cttctatccc agagaggcca aagtacagtg gaagggtggat aacgccctcc 540
aatcgggtaa ctcccaggag agtgtcacag agcaggacag caaggacagc acctacagcc 600
tcagcagcac cctgacgctg agcaaagcag actacagaaa acacaaactc tacgcctgcg 660
aagtcaccca tcagggcctg agctcgcccc tcacaaagag cttcaacagg ggagagtgtt 720
agaggagaaa gtgccccccc ctgctcctca gttccagcct gacccccctc catcctttgg 780
cctctgacct tttttccaca ggggacctac ccctattggg gtctccagb tcwtttttcc 840
cctccccccc ctctcctccc ttggbttaa ttatgytaak gttggrggar aatgaataaa 900
taaagtgawt ttttgmaaaa aaaaaaaaaa aaaaaaaaaa aaaaagaaaa aaaaaaaaaa 960
aaa

```

```

<210> 52
<211> 628
<212> DNA
<213> Homo sapien

```

```

<400> 52
cctttctctc tcctcatctt ggtgcatgtc ctttctgcag ctgcctttca gcacagggtg 60
ttccactggg ggcagctaac gctgagtgc aaggatggga agccacagg gcattttact 120
caagtcttct ctagtcaatg aggggcaccc agtgcttctg gggcaggctg ggtggtggtc 180
ccctaggtat cagcctctct tactgtactc tccgggaatg ttaacctttc tattttcagc 240
ctgtgccacc tgtctaggca agctggcttc cccattggcc cctgtgggtc cacagcagcg 300
tggtgcccc ccagggccac cgcttcttct ttgatcctct ttccttaaca gtgacttggg 360
cttgagtctg gcaaggaaac ttgcttttag cttcaccacc aaggagagag gttgacatga 420
cctccccgcc cctcaccaca ggtgggaaac agaggggatg tggtagagc caggttcctc 480
tgccctctc cagggtgttt tccactagtc actactgtct tctccttgta gctaatcaat 540
caatattctt cccttgctg tgggcagtgg agagtgtgc tgggtgtacg ctgcacctgc 600
ccactgagtt ggggaaagag gataatca
628

```

```

<210> 53
<211> 598
<212> DNA
<213> Homo sapien

```

```

<400> 53
ctccctccct ttttccacc ctccacctct caccggccct gcccagccc ctaagccccg 60
tccctaacc cccagtcctc actgtcctgt tttcttaggc actgagggat ctgagttagt 120
ctgggatgga caggaaaggg caaagtaggg cgtgtggttt ccctgcccct gtccggaccg 180
ccgatccag gtgcgtgtgt ctctgtctct cctagcccc ytytcacaca tcacattccc 240
atggtggcct caaraaaggc ccggaagcgc caggctggag ataacagcct yttgcccgtc 300
ggccctgcgt cggccctggg gtaccatgtg gccacaactg ctgtggcccc ctgtcccaa 360
racacttccc cttgtctccc tggttgccty tcttgcccc tgcctgaag cccagcgaca 420
caraagggg tggggcggt ytattgggar aaaggagagc aggtcaragg agggcatggg 480
ttggcagggt gggcgtttg ggccttyat gctggtttt cccccarag gacacaggca 540
gcttccaaa wawatttaty ttcttcacgg gaaaaaaaa aaaaaaaaa aaaaaaaa 598

```

```

<210> 54
<211> 592
<212> DNA
<213> Homo sapien

```

```

<400> 54

```

```

gtgttttacac agtaagatca agagtcaggc cctggaattt ccagaaccag cccgacatag      60
ttgaggactt gaaggacctg atcaccctgt atgctggaca agaaccctga gtcgaggatc      120
gtggtgccgk aaatcaagat mctggtraag accatratat gtaaacgctc ctttgggaac      180
ccattcgagg gcascggmg ggaggaacgc tactgtcag ycctggaaa cttgtcacr      240
aagcaaggca gygaagacaa cctccagggc wccgaccgc ccccggtggg ggaggaggaa      300
gtgctmttgt gagaggcagt ccctgctgtg aaagtgtgtg ggccccagc ccccggtcc      360
ccgcacgca tgcattccact gggcgaggag gaggccatgg agcccgagta gctgctgga      420
tcgctcgacc tcgcatgcgc gccgcgtcgc ctctgggggg atgctgcacc gcgtttccat      480
agcagcatgt cctacggaaa cccagcacgt gtgtagagcc tcgatcgtca tctgctggtt      540
atttgttttt tcctttgttg ttttaaaggg gacaaaaaaa aaaaaaaaaa aa      592

```

```

<210> 55
<211> 504
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (1) ... (504)
<223> n = A,T,C or G

```

```

<400> 55
agcgggacga gcgckngygg acttatcggg gcccggggac ccggggcgcg cggagggtgc      60
acttcaacct cctgcccgak cgctacgagc cactggagga gccggcgccg agcgagcagc      120
ccaggaagag gtaccggagg aagctgaaga agtactggca agaattgtcg gaaggtcatc      180
atmaakgat gccgctacgt ggtcatcggc ctgcaaggmt tcgctgcags ctactccgcc      240
ccgtttgchg tagccaccag cgtggtatct ttctgtcgct aatgggagct gctgtgycag      300
gtgccccag agtgaacggg agccccctgt gtgggaactt tgtgaatcct ggagcatctc      360
atacttgaac acacakcata tttggaagag aaaacatgcc tttctttgnt gaatcacatt      420
agtatgatga gtgagtcatn cctgcccac tcgtgagctt ctcacatctc tcagtcacac      480
gtggaccagc tgggtcaatcc tgca      504

```

```

<210> 56
<211> 749
<212> DNA
<213> Homo sapien

```

```

<400> 56
gtaagataca agtttcaggc aggggtctgaa ggactggatt ggccaaacat cagacctgtc      60
ttccaaggag accaagtctt ggctacatcc cagcctgttg ttacagtgca gacaggccat      120
gtgagccacc gctgccagca cagagcgtcc ttccccctgt agactagtgc cgtagggagt      180
acctgctgcc ccagctgact gtggccccct ccgtgatcca tccatctcca gggagcaaga      240
cagagacgca ggaatggaaa gcggagtccc taacaggatg aaagtcccc catcagttcc      300
cccagtacct ccaagcaagt agctttccac atttgtcaca gaaatcagag gagagatggt      360
gttgggagcc ctttggagaa cgccagtctc ccaggcccc tgcatctatc gaggttgcaa      420
tgtcacaacc tctctgatct tgtgctcagc atgattcttt aatagaagtt ttattttttc      480
gtgcactctg ctaatcatgt ggggtgagcca gtggaacagc gggagacctg tgctagtttt      540
acagattgcc tcctaattgac gcggctcaaa aggaaaccaa gtgttcagga gttgtttctg      600
accactgat ctctactacc acaaggaaaa tagtttagga gaaaccagct tttactgttt      660
ttgaaaaatt acagcttcac cctgtcaagt taacaaggaa tgctgtgcc aataaaaggt      720
ttctccaaaa aaaaaaaaaa aaaaaaaaaa      749

```

```

<210> 57
<211> 673
<212> DNA

```

&lt;213&gt; Homo sapien

&lt;400&gt; 57

```

agatgtcatg acccatcaac gtctcctctc attgtccctc tgctgggaca aagtgtttgg      60
tgtctggctg ggggacaacc aagagccccc aagtgcactt ccctaagggtc ctccagtgtc      120
tgaacatcag cgtgctaagt cagaaaaagg gtgaggatgc ttacccgaga cagatagatg      180
acaccatgtt ctgcgccggt gacaaagcag gtagagactc ctgccagggt gattctgggg      240
ggcctgtggt ctgcaatggc tccctgcagg gactcgtgtc ctggggagat tacccttgtg      300
cccggcccaa cagaccgggt gtctacacga acctctgcaa gttaccaag tggatccagg      360
aaaccatcca ggccaactcc tgagtcatcc caggactcag cacaccggca tccccacctg      420
ctgcagggac agccctgaca ctcttttcag acctcattc cttcccagag atgttgagaa      480
tgttcatctc tccagccctt gaccccatgt ctctggact cagggtctgc ttccccaca      540
ttgggctgac cgtgtctctc tagttgaacc ctgggaacaa tttccaaaac tgtccagggc      600
gggggttgcg tctcaatctc cctggggcac tttcatctc aagctcaggg cccatccctt      660
ctctgcagct ctg                                     673

```

&lt;210&gt; 58

&lt;211&gt; 994

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1)...(994)

&lt;223&gt; n = A,T,C or G

&lt;400&gt; 58

```

gacgggcccg gtgctgaagg gcagggaaca acttgatggg gctactttga actgcttttc      60
ttttctcctt ttgtcacaac gagtctcatg tctgatattt agacatgatg agcttttgtc      120
aaaaggggag ctggctactt ctgcgtctgc ttcateccac tattattttg gcacaacagg      180
aagctgttga aggaggatgt tcccatcttg gtcagtccta tgcggataga gatgtctgga      240
agccagaacc atgccaaata tgtgtctgtg actcaggatc cgttctctgc gatgacataa      300
tatgtgacga tcaagaatta gactgcccac acccagaaat tccatttgga gaatgttgtg      360
cagtttgccc acagcctcca actgctccta ctgcacctcc taatgggtcaa ggacctcaag      420
gccccaaagg agatccaggc cctcctggta ttctgggag aaatgggtgac cctgggtattc      480
caggacaacc aggggtccct ggttctctcg gccccctgg aatctgtgaa tcatgcccta      540
ctggctccta gaactattct cccagtatg attcatatga tgtcaagtct ggagtagcag      600
taggaggact cgcaggctat cctggaccag ctggccccc aggccctccc ggtccccctg      660
gtacatctgg tcatcctggt tcccctggat ctccaggata ccaaggaccc cctgggtgaac      720
ctgggcaage tggtccttca ggccctccag gacctcctgg tgctataggt ccatctggtc      780
ctgctggaaa agatggagaa tcaggtagac ccggacgacc tggagagcga ggattgcctg      840
gacctccagg tatcaaaggc ccagctggga tacctggatt ccctgggatg aaaggacaca      900
gaggcttcga tggacgaaat ggagaaaagg gtgaaacagg tgctcctgga ttaaagggtg      960
aaaaggtctt ccaggcgaaa aggagcnaca agtt                                     994

```

&lt;210&gt; 59

&lt;211&gt; 639

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 59

```

tgggtgggtg ggaggacctg cactctggcg cccagcagca ctggtecggc cactctgcgg      60
agatctccac gctggccctc agccacagtg cccaggctct ggctctgcc tggggccgaa      120
gcagcacgac cgccattgtt cagatcccg tctgggacgt gtctggcggc ctctgccagc      180
atctcatttt ccccatagc accaccgtgc tggccctggc cttctacca gatgacaggc      240

```

ttcttgtcac	actgggggac	cacgatggcc	gcacctcgc	cctgtggggc	acggccacct	300
atgacctcgt	gtcctccacc	cgctcccgg	agccggtgca	tgggtgggcc	ttcaaccct	360
gggacgccgg	cgagctcacc	tgtgtgggcc	agggcactgt	caccttctgg	ctccttcagc	420
agcgtggggc	agacatcagc	cttcaggtgc	gtcgagagcc	agtcccagag	gcagtggggg	480
ctggagagct	gacctcgtc	tgctacgggg	cacctcccct	gctctattgt	ggcaccagct	540
ctggccaggt	ctgtgtctgg	gacacgcgtg	cgggccgctg	cttcttgtcc	tgggaggcgg	600
atgacggtgg	cattgggctg	ttgctgttct	cgggttctc			639

&lt;210&gt; 60

&lt;211&gt; 470

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 60

ggcacgaggt	ttttattctt	tgcacaattg	tttcattggt	tgacacttaa	tgcactcgtc	60
atttgcatat	gacagtagca	ttctgaccac	acttgtacgc	tgtaacctca	tctacttctg	120
atgtttttta	aaaatgactt	ttaacaagga	gagggaaaag	aaaccacta	aattttgctt	180
tgtttccttg	aagaatgtgg	caacactggt	ttgtgatttt	atttgtgcag	gtcatgcaca	240
cagtttttgat	aaagggcagt	aacaagtatt	ggggcctatt	tttttttttt	tccacaaggc	300
attctctaaa	gctatgtgaa	attttctctg	cacctctgta	cagagaatac	acctgcccct	360
gtatatcctt	ttttcccctc	ccctcccctc	cagtgggtact	tctactaaat	tggtgtcttg	420
ttttttattt	tttaataaaa	ctgacaaatg	acaaaaaaaa	aaaaaaaaaa		470

&lt;210&gt; 61

&lt;211&gt; 535

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 61

ggcacgaggg	gttcttgggc	caatttcttg	ctgactcagc	cccactctgt	gtcggagtct	60
ccggggaaga	cggtaacat	ctcctgcacc	ggcaacagag	gcagcattgc	cgccaactat	120
gtgcagtggg	accagcagcg	cccgggcagt	gccccacca	ctgtgatcta	tgaagataac	180
caaagaccct	ctgggggtccc	tgatcgggtc	tctggctcca	tcgacagctc	ctccaactca	240
gcctcccctc	ctatctctgg	actgaagcct	gaggacgagg	ctgactacta	ctgtcagtct	300
tatgagagca	gcaatcattg	ggtgttcggc	ggagggacca	agctgaccgt	cctagggtcag	360
cccaaggctg	ccccctcggg	cactctgttc	ccaccctcct	ctgaggagct	tcaagccaac	420
aaggccacac	tgggtgtgtc	cataagtgc	ttctaccctg	gagccgtgac	agtggcctgg	480
aaggcagata	gcagccccgt	caaggcggga	gtggagacca	ccacaccctc	caaac	535

&lt;210&gt; 62

&lt;211&gt; 696

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1)...(696)

&lt;223&gt; n = A,T,C or G

&lt;400&gt; 62

gccagtatcc	aatggggaat	cagctctaaa	aggaaccaga	ccaacgtttt	ccagcccntt	60
catntnggtg	actgagggga	ggaaagaatg	ggagggggta	ttcttgtcta	gtggatggaa	120
aggaaacaca	ctgtcaaatt	actatatctc	cttggttttc	tattacagta	gaattctcca	180
gccatatatt	tattgtctat	gggggaagtt	ggagatgggtg	accttgatta	gaagtgtctg	240
gaggggggata	aatggagggg	ataagattca	gttggttttg	gaaaatgtta	aagtcctaaa	300



```

ataatgcgtc catctgaaga attttttcta aaaccagagt ttataaaaat atcactgata 360
cagcctgccc cctcatttcc ctgccacagg agatgtcttg gactagagac acttggttaa 420
taatagcttg tctctgatat tcccagtagc ttcctctgtg gtgaggaaag gatagaaatg 480
ttcaggacat catcatacag gctcctcacc taaaaagttc cagtagcagt gacgcctaca 540
cggaagactt ggaactgcaa acaggctggg gtcacctcag tgacatctga cgctgtccaa 600
ccagaagtgc gatttttgtt ctgggggtga aggaggaaac agacgtgtac taaaggacta 660
aaataatttg tctatacaaa aaaaaaaaaa aaaaaa 696

```

```

<210> 63
<211> 256
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(256)
<223> n = A,T,C or G

```

```

<400> 63
taatatattat tgaagnccta ctttgtcctc tacatagggt agcttctgtc agggaatgtt 60
ggttcttccc aagaaacact gattttcttt caggagact tcatgtgttc atttatttcc 120
accacagcag attttaagaa attataatat gtaatatattg atatctataa agagtanac 180
taactgtaat aaattatgaa gcatactaata gactacatat gacctatnac acatatacat 240
taaaacattt taaata 256

```

```

<210> 64
<211> 678
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(678)
<223> n = A,T,C or G

```

```

<400> 64
ggcacgaggg ctanaagcca tcgcatcttt ctccgtattc aacaagccca ttataacaag 60
tctttcacgc tcaaccattc aaacaatccc aacaatccca gcctaagaat ttcttggtcat 120
ttgatgaaaa tggtctgtgg ttttgctatc tttaagctcc ctgggaagca ggatataagc 180
ccagggtctg caggtctctt ctagctacct gctcttgac atagagtttg cctcatggtt 240
gcaagaaggc tgccatagct ccagatggca catagacatt ccaggcagca ggaaaaagga 300
aggagcactt aaccgaagga ggtcaggagg ttggttagtc tccacctgaa gaagagagcc 360
atgaaccagc ttcagttgac taacgggctc ctgtgagtgc atctttggac ttttctggag 420
gttgaatatc agatgtggta tgtgtcttaa agcagccacc aactcctccc attaccttcc 480
aagtgaagcc aactacacga tggagcctct tccctgccct tggatctggg ctgagcctct 540
gacttgcggt gaccaacaga atgcagtga agtgatgccg atactacct cctgcccta 600
gacttgggat acctgcagct atattcttcc attcctcagg acttgcaaaa acgggtctca 660
ncatgccttn ccagagcc 678

```

```

<210> 65
<211> 678
<212> DNA
<213> Homo sapien

<220>

```

<221> misc\_feature  
 <222> (1)...(678)  
 <223> n = A,T,C or G

<400> 65

```

ggcacgagga ggacacactc aggatagaaa ctttatgtgt atagggaatg tggtagagcc      60
tttagccagg agtcatactt catcagacac cagaggacac acacagtgtt gtggcttttt      120
cagccattgc tagataccaa agtggagaca ttctgtgtgt gattatgcat gagactgtac      180
tggtaagact tgtatctcca tccacctgaa ggagaattgc tggctcattt tcaggagccc      240
tgcccttcct cactgtggat ggtgggttgt ggaaacccgg tcaggtaatg atagtggcag      300
gaggcagtca aatgcccagg cagatagggg tgggtacctg gtgaaaccca acctaaagc      360
tgaagacagt ccgggctaaa tcctcatact gaattgagaa cctgtcttcc catttggtgt      420
gctttcctcc gattgatccc aaccttcac ctattttacg tatacctgcc ctttcctaata      480
tggtttttac actgctgtgc ccaccttttg agtgggtgcct ttgcatactt acaaatacagt      540
caacgtgtat tcccctattc tgagcccata aaagacccan actcagctgc agtgaggaga      600
gaaatcccct gctgnggggg gtggggacca ctccctgcat ccctctncac tganagctgt      660
ctttttgctc aataaaat

```

<210> 66  
 <211> 606  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(606)  
 <223> n = A,T,C or G

<400> 66

```

gcatcaactg ccgccagccc ttcattntnt ccgcctcttc ctagacgtgc tacacctggt      60
tgagttctac ctggaggaag ggatcactga tgaagaagcc atctccctca tcgacctgga      120
ggtgctgaga cccaagcggg atgacagaca gctagagatt gcaaacaaca gctcccagat      180
tctgcggcta gtggagacca aggactccat cggagatgag gacctgttca cagctaagct      240
gagctttgag caaggtggct cagagtccgt gccagtgggt gtgagccggc tgggtgctgcg      300
ctccatgagc cgcggggatg tcctcatcaa gcgatggccc ccacctctga ggtggcaata      360
cttcgctca ctgctgcctg acgcctccat taccatgtgc ccctcctgct tccagatgtt      420
ccattntgag gactatgagt tgctggtgct tcagcatggc tgctgcccct actgccgcag      480
gtgcaaggat gacctgggc catgaccagc atcctgggga cggcctgcac cctctgcccg      540
ccttgggggtc tgctgggctg tgaaggagaa taaagagtta aactgtcaaa aaaaaaaaaa      600
aaaaaa

```

<210> 67  
 <211> 579  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(579)  
 <223> n = A,T,C or G

<400> 67

```

ggcacnaggc ccaattatat gttcattttg tatatttttt ggtcggggga aaaattgacc      60
tgcagtanaa aaacctttga ccatttttat gtccattgga tactttcctt tttatcatct      120
taaaaaaaga taactagtac taatcattgt agtggcctaa gtgtgattta actcttgaag      180

```

tcacaccctc	cgaagatga	gtagaaacca	gcaccagcac	agcccagatc	ttctctttcc	240
tctccttttc	ctcattttatt	cctaaaggaa	tctgaccatt	ttacgtctct	acggcccaaa	300
aaaagacaaa	aataaaaatt	ccttttttatt	cctgtcaact	ggatggaaac	acaaatttca	360
tggagctgtg	taccatcgaà	gaaacctggt	gtctggcatg	aaattactgt	aaagaacttc	420
ctgtaaaaca	cgttctttta	caaaactgaaa	tgaaaagcat	tggagcgtct	gaatgaaaga	480
cgtgacctcc	tgctgggact	ctgatggtct	tcagcattca	ccttcgtgtg	tcttcagtgt	540
ctcattgtca	tccttgcttc	tgtttggtct	tagagtgtt			579

<210> 68  
 <211> 258  
 <212> DNA  
 <213> Homo sapien  
  
 <220>  
 <221> misc\_feature  
 <222> (1) ... (258)  
 <223> n = A,T,C or G

ggcatntggc	cctgaccana	tccttaaccc	tagcgtggc	ttgggagatg	gggggttggg	60
tagcattntc	tttcttgcc	cttccttatt	ctaggaaaan	atgggttcctc	tccttgtgtg	120
tgtctcttcc	ccccaccct	aattcttctg	ctctgtttgg	gaagacgtgg	aggaaaaggt	180
gacttctgcc	cccaccgctc	ttacccccac	tgtantggcc	tttggagatg	ccccacctc	240
ccccccacca	actctcgc					258

<210> 69  
 <211> 628  
 <212> DNA  
 <213> Homo sapien  
  
 <220>  
 <221> misc\_feature  
 <222> (1) ... (628)  
 <223> n = A,T,C or G

ggcacnagga	gaattntttg	gaataaaaact	gactgataaa	actggaaatg	ttgctcgggc	60
tgtaactgg	caggaaagat	ttcagcatct	gaatgagtcc	cngcacaact	atttaagaat	120
cactcgtatt	cttaaaaagcc	ttggtgagct	tgatattgaa	agttttaaat	ctcctcttgt	180
aaaattttatt	cttcatgaag	ctcttgtgga	gaatactatt	cccaatatta	agcagagtgc	240
tctagagtat	tttgtttata	caattagaga	cagaagagaa	aggagaaagc	tcctgcggtt	300
cgcccagaaa	cactacacgc	cttcagagaa	ctttatctgg	ggaccgcctc	gaaaagaaca	360
gtcggaggga	agcaaagccc	agaaaatgtc	ttccccctctc	gcctccagtc	ataacagtca	420
aacttctatg	cacaaaaaag	ccaaggactc	caaaaattcc	tcctcagctg	ttcattttaa	480
tagcaaaaca	gctgaagaca	aaaaagtggc	accaaaagag	cctgtggaag	agacagacag	540
gccccccan	agcccagcaa	tgaagctgcc	aagccaagaa	atacagaaga	aggacagtaa	600
tgctganaac	atgaattctc	aacctgag				628

<210> 70  
 <211> 439  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature

<222> (1)...(439)  
 <223> n = A,T,C or G

<400> 70  
 ggacacgaggc gcacacccgc tgggcgggccc ctgtggagac cctggaaaac atcategccca 60  
 ctgtagacac gaggtgcct gagttctcag agctgcaggg ctgtttccgg gaggagctca 120  
 tggaggcctt gcacctgcac ctggtgaagg agtacatcat ccaactcagc aaggggcgcc 180  
 tggtcctcaa gacggccgag cagcagcagc agctggctgg gtacatcctg gccaatgctg 240  
 acaccatecca gcacttctgc acccagcacg gctccccggc gacctggctg cagcctgctc 300  
 tccctacgct ggccgagatc attcgctgc aggaccccag tgccatcaag attgaggtgg 360  
 ccacttatgc cactgctac cctgacttca gcaaaggcca cctganctg atcctggcca 420  
 tcaaggggaa cctatccaa 439

<210> 71  
 <211> 328  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(328)  
 <223> n = A,T,C or G

<400> 71  
 tnggaactca ggtttacnca aactgagggg gccccagccc tngtacncc cctgttacc 60  
 caggatccat ntgcctcan aaaagngttc aggtacagca gctgaggctg ccctgaggaa 120  
 tcaaggggcc attaccaagg ggcaggaaaa ggatatgtaa naggnngcct tcatggtana 180  
 gcttgaccca anaactactc cncattngga tggcccagac tgactccatc ccctgacttt 240  
 ccctttgact tcnccttggt tgtaataaaa acaataaaat ggaaggtgct gtggacggga 300  
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 328

<210> 72  
 <211> 721  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(721)  
 <223> n = A,T,C or G

<400> 72  
 ggtttctggt gctattctgc caagtatatc agctcctcct catgtttcaa cattccatct 60  
 tcccgtttct atcctcgact ccaaagtaag ccttcttagc tccaatcagg gatgaggggc 120  
 tcaacctctt ctgtcctcaa agaggccaaa cgagtgcca cagtcggtag ccttcaacttt 180  
 tagatgtcct attcatgtaa aaaagaagggt gccccacca ggcttacatc agcaataagc 240  
 aattctaatt caacgatggt gtccacattt taccacagtg tgtgcccatt tatgcctttg 300  
 tgcccgtgta attattgtta gcgccccttt cacttagagg ggtgatgata aactgtggcc 360  
 accttgatta caaccacat ttctgcttt ggggagcttc caagtaacag gccatttctt 420  
 acctccctcc aggaacagtg ggcactgccc accacctcgt gtctgctcat aggatgacgc 480  
 tggagatccc cacacttact ctaccctctt ggcaaattgg cattccgggt gtggtttttg 540  
 tttcctttta cacattaaat aaatgagtat ataggatgtg aggggagggg tgagaacaac 600  
 tagctgtagc atgtgtaggc tatatacttt accattcgac ttcttttctt tttttttttt 660  
 aaaaaaaaaa aagtgcctga ctgggttcaa gcttcatcat gaaaaaaaaa aaaaaaaaaa 720  
 a 721

<210> 73  
 <211> 596  
 <212> DNA  
 <213> Homo sapien  
  
 <220>  
 <221> misc\_feature  
 <222> (1)...(596)  
 <223> n = A,T,C or G

<400> 73  
 attgtacaga acagcccagg tagggaggac agntgcccga ggtcccatag gantgcatgc 60  
 ctcaagccca cgtcatgcag agccactcag ctcaccctgn tcagggcacg tggtttacct 120  
 gcattcccct nttgcaggtc ctacgtgtgg aggacgtacc atntgacctn tgcaggagag 180  
 aaactcacgg aagacagaaa gaagctccga gactacggca tccggaatcg anacgagggtt 240  
 tccttcatca aaaagctgag gcaaaagtga gcctccagac aggacaaccc tnttcatcac 300  
 tgggtggctga gctttttccc agcaggaatg ggtcctcgaa tcatcgtgcc tntttcacan 360  
 aaaggacgtt gtggtggcct caccacaggc atgcccaca gtaactgtca gcataaacct 420  
 gggggccctc aggactagga cagggtgagc cagtgtctcc tcctttcatg tacttggcct 480  
 gagactgacc tctcctagg tccaaatgcc ctagtacat ggagagccca cggcctggcc 540  
 cactgtataa aataaacctg tttgcttntt agtttgaaaa aaaaaaaaaa aaaaaa 596

<210> 74  
 <211> 302  
 <212> DNA  
 <213> Homo sapien

<400> 74  
 ggcacgaggc ttaaaccaag tttctctgca gctctttcgg ttctgcttac agtgtgtggg 60  
 aaatctgatt tttttcccct agtaatagtt tgataagaaa tttagtgtat tgactgcctc 120  
 agtgacacaa tttatcttta aagggtgtgga agctgggtggg gaccaaagtgt tacctgtgtt 180  
 tttgctgttg attgctatct tcagaagcaa accatgtttt tcaattacag taggagtcaa 240  
 caaatttgagg attttagaag ggggaggagg gagctatttg tgtaagactg ctgtcatatt 300  
 tg 302

<210> 75  
 <211> 635  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(635)  
 <223> n = A,T,C or G

<400> 75  
 ggcacgaggc ttaatgtgac tttcctgagt taacactgaa tagtatttct aaaatttttt 60  
 gtacttttatt ttttaatgta acttggttcta tctatctata tatatatttg atagtttgtg 120  
 gaataatata ccccagtatt ttccatatta aatgctaatt atcttttgat ttctttttca 180  
 taagcagatc tggcatttat tacagggctg ccgcttaaga gaactcatta taatgaacgt 240  
 ttattatatt ttgcagttcc atgcctgttg tccattgatt gacatgagca cccctgtttt 300  
 ctctggagaa atacctcccc tctctggggg gcttcctgtg gtagtgtctt tcagggtatcc 360  
 gttccactag ctacaggtga gcattttacc cattgttggg taatggtaat ctctttttca 420  
 gaattttgag tctgtaattc atttgtacat gaaccagaaa atgtgggaac tcattcattc 480

ttgtcccaga attctgttga gaacatccat tcattctggc taattgatta caagaataac	540
tgnggatacg atccctttan aacctgcttc tctgatctgn gtgtttcctc acttctcaat	600
aaaaatgtct tttgctaaaa aaaaaaaaaa aaaaa	635

<210> 76  
 <211> 678  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(678)  
 <223> n = A,T,C or G

<400> 76	
ggcacgaggg tcttgattct tggcttgccct ctccctccaat tccaaactta gtgaaatggc	60
cttaagcatt ttaaactgta tgtatacatt agcgcattca tgcctttcta aacgcatttc	120
aaatgtcaac caggaaggca caccactgta ttagttttat actgccgtg taaaatttac	180
cacaaactta gtgacttaac acaaatttat tgcaattctg taggctggaa gtctgactat	240
gggtctcact ggactagaat caaggctggc aggctgcctt ccttcctgga ggttctaggg	300
gagactctgt ctccctgctcc ttcaggctgc tggcagaatc cacatccttt cgggtggcagg	360
gccaaaggtcc ccactttctt gctgactgta aactaaggcc acttccagct tgtagaggct	420
gcctacattc cttggctctt ggccccctcc tccatcttca gagctagcag gttcagctctg	480
tgtcacgaac catttctctg gttccctgca gacaggaaag gttgtcccta aggactcatg	540
agattagggtt gggcccagcc agataatata tgataatctc cctcctcaag gnttttaata	600
taaacacat ctgcaggaca cattttgcca tgtaactaac attcactggg ccaggggatt	660
aaggaatgaa ccctcttt	678

<210> 77  
 <211> 669  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(669)  
 <223> n = A,T,C or G

<400> 77	
ggcacgaggg agaattcttaa aaaaaaaaaa acgtttctca ctgtcttaaa tagaattttt	60
aaatagtata tattcagtgg cattttggag aacaaagtga atttacttcg acttcttaaa	120
tttttgtaaa agactataag ttttagacatc tttctcattc aaattttaaag atatctttct	180
cctcttgatc aatctatcaa tattgataga agtcacacta gtatatacca ttttaatacat	240
ttacactttc ttattttaaga agatattgaa tgcaaaataa ttgacatata gaactttaca	300
aacatatgtc caaggactct aaattgagac tcttccacat gtacaatctc atcatcctga	360
agcctataat gaagaaaaag atctagaaac tgagttgtgg agctgactct aatcaaatgt	420
gatgattgga attagaccat ttggcctttg aactttcata ggaaaaatga cccaacattt	480
cttagcatga gctacctcat ctctagaagc tgggatggac ttactattct tggttatatt	540
ttanatactg aangnggcta tgcttctgtt attattccaa gactggagat aggcagggct	600
aaaaagggat tattattttc ctttaatgat ggggctaaaa ttcttctat aaaattcctt	660
aaaaataagg	669

<210> 78  
 <211> 134  
 <212> DNA

<213> Homo sapien

<220>

<221> misc\_feature

<222> (1)...(134)

<223> n = A,T,C or G

<400> 78

ggcacgaggg gtcgatttta atagcgaatc ctttttcttg tagaggtaag taaaatcttc	60
ctgacaaggt tgcctctnt tcacggcaca gacaatgggc ggnctgttta tgaggggtga	120
gaagngacnc ccgc	134

<210> 79

<211> 396

<212> DNA

<213> Homo sapien

<220>

<221> misc\_feature

<222> (1)...(396)

<223> n = A,T,C or G

<400> 79

tanatgcaag tntnttgtn gataacgta ttgagatatt acncctagtc tgtggcttga	60
ctgttttctt tatgtctttt gatgaatana agttttaaat ttgacaagg tcaaatttat	120
ttttttcttt tgtttgatat tttttntctc caatttaacc ccaagatttc anatattntg	180
ctctattana taaactttat atttttatat ttgtgatctn ccttgaattg atatgtatgt	240
tgtgaattat ggatcagggg tntttttttc cccatacaa gtatccagtc attgtaacnc	300
tgtttattga aanaattatc ctttcctcat taaattncct tgccaattag taaaaaatca	360
attaccata aaaaaaaaaa aaaaaaaaaa aaaaaa	396

<210> 80

<211> 731

<212> DNA

<213> Homo sapien

<400> 80

tctacatcat cctgagagcg ccttcagact ggacagaact gtaggaactg ccgtctctta	60
tggtagatgt ggcccatcac aatatacctg caggtttaaa ctctggacac ggcttattgc	120
agctggagga gtccagggtt aacatgtgga ctcggaagaa gaagccgctg ggggtgaggt	180
acaggcggtt catcttgtag agcccgctcc gatccaccag cagggtcacc agccggatgc	240
ccgcgcccag cacgtccaca tcgtgcacga tcccgttcac cgcgaattcg ctctcgcaga	300
acgcctcgcg ttcgtccaga tcggaccggc aggcgcgcgc gcagggtca gcgcgggct	360
ccgcgtccct ctgcggcggg ctgaggccga agctgaggct gaagcgcggc ggcggcgcg	420
gcgcgggtcc cggggccagc gcggcggcgc gcggccggtt ggggtggcg tggccggcg	480
agccctcgtt ggcgagcgcg ggcggccggg cggcggggaa gcgcagggcc cgtgagcgcg	540
cgcgtcgcgg ggcgtcctcc gacgcgggct cgcggcgcg cgacgcgcgg ttctcggctt	600
gcgccaggcg cgggcccggc gggccagcgg gcgcgcgtgc ggcctcccgg aagctggcct	660
ggttgctggc cggcggtggg ggcttcgggc ggcgcgggtc aggccaggcg cggacggcg	720
gagcgacctc g	731

<210> 81

<211> 396

<212> DNA

<213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(396)  
 <223> n = A,T,C or G

<400> 81  
 cccttcatcc caatccntga gaagttcttt ttctgcttcc tgcccaacat natcaacagg 60  
 acctatttcc cattttctctg ntcttncntn anccagntat tggngngngt tttgaaatgg 120  
 ctnatnatga ggaanagtnt gatccnacac ttggagganc naggggtgca ggtggtnntt 180  
 tgggtgcctta atgaanagtc ggattttgaa gcagccttna gcgtgggagc cacnggcgtc 240  
 ataacggatt atcccnagc cctnnggcan tacctggaca accatggnc agctgcccg 300  
 acctcctaag tccagaancc tngaggtntc ctgtttntnt tccatgaaaa ataaatattt 360  
 gcctttcgat caaaaaaaaa aaaaaaaaaa aaaaaa 396

<210> 82  
 <211> 502  
 <212> DNA  
 <213> Homo sapien

<400> 82  
 ctcagaggaa agaagaaaag ggccaggagt cagacgtcac acccgggggc tcccctttcc 60  
 catgtagaag tggtggcatg cgtcagtctc ctttacagag ggggtggatgt atccgtggag 120  
 gaggggcctt ctctctttct aattgcacta tacttgttct agcttcagtc tggagatact 180  
 taagacctcc atggggtcgt gatccataga cttagcaagt cttgccttat ctatggagct 240  
 cgatgggtgag aattgtgacc attgtctgat gtccatagtt ccttccccct agattgtttc 300  
 tttccacgga ttgtgttcat ctgaaccatt ttatttttta ttaccaaag tactgtactt 360  
 ggctatttgc agtgttttca aaaccaaattg tttctttttt tgtgttttta atcttcgata 420  
 cttgggtgcaa tagaagctgc aaagatgtgc cactttatct atgaaatgga gttttgtata 480  
 ccaataaatt ctagttttaa aa 502

<210> 83  
 <211> 666  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(666)  
 <223> n = A,T,C or G

<400> 83  
 gatttttggg atttgtactt atttatgaac ttacttgag acagaatatg gttaaaatta 60  
 ggaacatcta ctttgaatga ggtttatttt tctattttga atttgcctta tgtatattca 120  
 aaggcttatg gaaatactgt aaaggaacat taggaaaagg acaaataaggc tataaccatc 180  
 tatcttaaaa tcagaccctt agtataagca cctctttttc ttttctctt tgacaattta 240  
 gtctcttatt taggtccatg taattaattt cattccatta ttttttagct gtttattcta 300  
 aaaaacaaaa attttcagcc actcccattt attctccat gacatgggtc tatatagcag 360  
 tacttaacac agtgccttgc acatagtagg cattcagtaa atacttacat gcatgaatga 420  
 ataatgtatt ttcagtgtaa caaatttatt ataaaagtgt agttcgactc ttcttggctc 480  
 tggagtttga gagtacagaa ttacagggaa tgaagagagg tataagtaga tattttaatg 540  
 gaaatgaagt ataaattaat aacttgactt accctccat aagttactgc taactggaga 600  
 tacccttgta tgcccaacct gtaaaggaaa aagntctatt tctatattta taggacattc 660  
 ttctca 666



<210> 84  
 <211> 199  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(199)  
 <223> n = A,T,C or G

<400> 84  
 aanntactta agttattcaa agaatgttat ctttcttgca agagtaatth aagcacatgg 60  
 gaaagattct agactttttg tttcttgcaa canacagtgc cctctgctgc tagaaacctt 120  
 tttctactta ctatcatttt tattgtggct tgagctcanc tcaatctggt gcagatgacg 180  
 ctggacaact actaaccaa 199

<210> 85  
 <211> 670  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(670)  
 <223> n = A,T,C or G

<400> 85  
 cacttcagga gagaacttca tagcacaatg tctttctata agatattttt aatgatttag 60  
 tattttacaa catttggtta ccatattttg atataccatt tttttctatc tgcccagttt 120  
 tattaaaaaa actatatatt attttctaaa gaaacaatca tatttttata caaaattatg 180  
 ttttcaggta acgaaataga tgtagggtac agtggaaacat aagcagtggt acccctgggt 240  
 gggagtcagt attatacaac aaatgggtgag ctggaacatg ccctgtctgt gctgtccctc 300  
 ctgtgctggg tcgcggtatg gtaggcaaca ttgccttata acgctagggt cacctgacac 360  
 tttaaaagga aaaaaagttc catagagttc tgtggtcaca aaattgtttt gcttttatca 420  
 aatactttta tagaaccaaa gttgcagata ttggaatgta tgggaagtatc tcagtctctg 480  
 cataagagga ttaaagtatg aaaggatcat ttaatgactg ttttacttat aagtcattaa 540  
 gtaatccacc atttcttatg gatgatgctt aagcctgggt aggtttgtac tctaaggagc 600  
 ccagatcata atgcagngca tttccttanc ccttagagtt tcttgcaaac atttaaaaaa 660  
 agacntattt 670

<210> 86  
 <211> 401  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(401)  
 <223> n = A,T,C or G

<400> 86  
 atcttccac gagggtgatt ctggccttca gagaccagga gggagtgctt gggccgcang 60  
 tgtggcactg tgggtgagag gtgtgtcttt gcacacacag tgcagcggga acggtggggc 120  
 tggctgggtc tgaagacaga cacactcctg agccaagggtc ttgtcttcaa cctccccgtc 180  
 ccgttgctcc attttgctct gtgaagggtc aaatcccttt ctcccttcc catctcaggc 240

```

tctcctgttt tccctcaggg tccagtatgc ctttgagctt tagctgttaa aaaggaaccc 300
ccgtgacttg acacagcttt cacanctggc tgctangacc ggcgggctgg gtgttcacgt 360
gtgtctgtgt catggatgca atgcangccc tgnangactg t 401

```

```

<210> 87
<211> 373
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(373)
<223> n = A,T,C or G

```

```

<400> 87
cttttttttt ctttttttta ttgaaaact actttancaa taattaattc catgattatc 60
acattctgcc attaagggat attagtaccg taatactgaa gaaattttat taagtctgaa 120
cttctggggt aggcagcttc tttgtttctt ttctatccac ccttgctcgt tgaggatatt 180
gtttcttgac taataaaccc tttgatactt taaccagaaa tcagctcata aagctatttt 240
tgagtatagt tngggnaaaa aaaaaagggt aacttgggna ataccttcca nctgacctc 300
cntntaccaa gatatttttc agggctttta tttactatgc nctaanacta tgcncttttt 360
ctgaaatatt ttn 373

```

```

<210> 88
<211> 507
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(507)
<223> n = A,T,C or G

```

```

<400> 88
ccaggagcag cccattcctg atgcttcttc agagactcct gcaggcagcc aggccacang 60
acccttggtg tcccacccca cacacgccag attctttcct gaggtctgggc tcccttccca 120
cctctctcac tccttgaaaa cactgttctc tgccctccaa gaccttctcc ttcacctttg 180
tccccaccgc agacaggacc agggatttcc atgatgtttt ccatgagtcc cctgtttgtt 240
tctgaaaggg acgctaccgg ggaagggggc tgggacatgg gaaaggggaa gttgtaggca 300
taaagtcagg ggttcccttt tttggctgct gaaggctcga gcatgcctgg atggggctgc 360
accggctggc ctggcccttc agggctccctg gtggcagctc acctctccct tggattgtcc 420
ccgacccttg ccgtctacct gaggggcctc ttatgggctg ggttctaccc aggtgctagg 480
aacactcctt cacagatggg tgcttgg 507

```

```

<210> 89
<211> 796
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(796)
<223> n = A,T,C or G

```

```

<400> 89

```

```

gactttaatg ggnatatggg ctctacactc atagggaccc acatgttcca agcctccagc      60
tgtgtgttct gcttcctcaa gcagcagctg actctgatat tccccatctc accttagcag      120
tctatatcaa gcaaagggca tgcagaatcc catgcaccca tcatgtatgt aataaaaggc      180
agatattatg tggctctctg aaccagtctt aggcattggag gttgaggatc aggagtgact      240
tgagggtact gactggcaga gcaggagccc cgttattttg gacaaacacc gccactttaa      300
gtttcagctt catttttagc cttctggatt taaggaaatt acttttttaa aaactataag      360
cagccaaaaa aagcagacag taaaatgcag ataaaacagc tcgggcacag aggaagggtg      420
aggaaaagtc tcttgggtaa ctgccaaact tcaccctcat acaatgggcc ccagtaaaac      480
agtgggcctt aataagcaca ttcctttccc tccaggtgca ctaaaatagg gaagctaaaa      540
gcagacttgg ggggtatgcc tacagctgca gaaaaatgta taaaagcaaa cacaccactc      600
tccctcccat ataagcacia caaaaaaaca cagaagcagt ccaagcctnt aagaaactct      660
cccaccctaa atccttaaac actcttagtc tgtagaaaag actctaactt aattcagcca      720
gcagcccctc tcaggtgtgt tttntntaaa ataaacctgt ntaaccatc aagccaaaaa      780
aaaaaaaaaa aaaaaa

```

```

<210> 90
<211> 462
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1) ... (462)
<223> n = A,T,C or G

```

```

<400> 90
tagaacatnc tgaatttttt ttgtactgnt ggactntatt cagtgtcatg tcctatatct      60
gatcaagttt tcaagnagat aattttanaa tgaaaaagaa aatcctcttg tnggaaacaa      120
aagacgtttt atatgtgcag tatgacaaan aggagtttca nagacaactt tgaatccttg      180
tcagcctgga gaccagcnc agaggaatnn ccaaggcaaa ctcccatata tttgttcccc      240
ccaaattgct gccctacag actcaaagct ctttttcttt gttttgttgt ttntctaaaa      300
atttactggt ntttgtcgat gctatataag ccagggagtt ttaagacgcc agctntttga      360
natttgntca ttccctgta tttcccat anatattaca tatacccgng taataaattt      420
atgtttgtta aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aa

```

```

<210> 91
<211> 591
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1) ... (591)
<223> n = A,T,C or G

```

```

<400> 91
cgggccttgg aagcctttga cctggatcct gcacagtggg gagtcaatgt ccagccctac      60
tccgggtccc cagccaacct ggccgtctac acagcccttc tgcaacctca cgaccggatc      120
atggggctgg acctgcccga tgggggccat ctacccacg gctacatgtc tgacgtcaag      180
cggatatcag ccacgtccat cttcttcgag tctatgccct ataagctcaa ccccaaaact      240
ggcctcattg actacaacca gctggcactg actgctcgac ttttccggcc acggctcatc      300
atagctggga ccagcgcccta tgctcgctc attgactacg cccgcatgag agagggtgtg      360
gatgaagtca aagcacacac gctggcagac atggcccaca tcagtggcct ggtgggtgcc      420
aagggtatcc cctcgccctt caagcacgca gacatcgta ccaccactac tcacaagact      480
cttcgagggg ccaggtcagg gctcatcttc taccggaaag gggngaaggc tgtggacccc      540

```

aagactggcc gggagatccc ttacacattt gaggaccgaa tcaactttgc c 591

<210> 92  
 <211> 647  
 <212> DNA  
 <213> Homo sapien  
 <220>  
 <221> misc\_feature  
 <222> (1)...(647)  
 <223> n = A,T,C or G

<400> 92  
 gaataactag aaatttattg gatcaggttt cacatttgca tttttgaaaa ctactaccaa 60  
 aaagattttca ccaatttaca actccatcat tagtaagaat gcctgtttgc ctatagtctg 120  
 ccaaccctga atccttaaaa atttttgcca atctggtagg caaaatttct ttcttttctt 180  
 tgaatattaa tgaggaggaa catcttttca tgtttcttgg ccatttgcac ttctatttat 240  
 gaattgcttt tgcccatttt ccttttttta attatgaaag tctaatagact accttctcat 300  
 tgtataaaaa acacagttct ttgaatagag agaccctttt ctccaatgct accaatcaca 360  
 ttccacttac cacagtttaa catacatcct ctatgcacct ttccgtacga atatacatatc 420  
 acataaaaaac actttttaca taaataggat ctcatattct gtagcttttt aaaatttttg 480  
 nctcaaaaaa aganaacang gctttaaatt tctttaatgg gtgaatatga ataaatacta 540  
 tgaaaatgcc attatttatt cccttaattt ttttctcttc gctattacat tgccaaagna 600  
 aacatcctat tcanatgtct ttngncatgt gngngaatat ttcttta 647

<210> 93  
 <211> 740  
 <212> DNA  
 <213> Homo sapien  
 <220>  
 <221> misc\_feature  
 <222> (1)...(740)  
 <223> n = A,T,C or G

<400> 93  
 gtcattgatag aggaaattaa tgcaaggaaa gaaaacaagc ccagttgggtt tggcttgtgc 60  
 tactgcaagg ctttacaatc agattatatc acatacatag atgaactcct gacctatatac 120  
 aatgcaaaac ccaacctgtt ctctatgtct ctaacggatc cacatctggc tctgaccgtc 180  
 ttctttggcc catgctcacc ataccagttc cgtttgactg gccaggaaa atgggaagga 240  
 gccagaaatg ccatcatgac ccagtgggac cgaacattca aggtcatcaa agctcgagtt 300  
 gtacaagagt ctccatctcc ctttgaaagt tttcttaaag tctttagctt tctggctttg 360  
 cttgtggcta tttttctgat ttctctataa gtaaaagatc tcctaaatgg aagatgcaca 420  
 gagtagattt acaatgctcc aatctctctc ttacagcaat attgccttca cagttataaa 480  
 ctgtattcaa atagttaaagg ccaccctctc gcttccctgg ctggccccag ggctaccact 540  
 ggtattcctg agcctctccc agctccactt ctaatgctag agaatgataa ctaagacttc 600  
 tgtgcatttg aaggttggtg gaaagttaca gggtcatttt anaaagaaaag ctgttcttga 660  
 cagcactctt gagccatcat acctcttcc ntataaacta ttttccagaa ctcnactaaa 720  
 accccttact tcacaaatga 740

<210> 94  
 <211> 608  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(608)  
 <223> n = A,T,C or G

<400> 94  
 ttttacaatc ctaggaaggc ccaccaatct catttcacgc gccagggcgg ctgcagttgg 60  
 aggccgaggg cagccctctg ctactgaat gtcttgcatg tgctgactgc tgccgcagct 120  
 gctgaacatg cccaccgccc caggcccagc actgcttggt gggtcagcat ctagtgtctgc 180  
 tgtcacatct ttgtctgcac agccagtagg attgcctcag ccaggggggt tatcagaagg 240  
 tgtgcaaggc ctttggggga actgagcccc tatagtgggc agtctccttt accttcccac 300  
 ctccctgaaa agcacagaag acagtgcctt ggtttgtgtt ttgaagcaaa caagtcagct 360  
 ttctggcttt gccccaaaac tgtgatggaa cataataaaa ctggagatat ggtttttaac 420  
 actgcaaaaa ggaaaaagca tcaagtttct acttctggct ggaaagcaaa accaatctca 480  
 gctgacaagg ctgggcaaac taagttttcc tgagcccatt ttcctttgag ccctgacctt 540  
 ncctgcctta cctcatatag gtttgggtta agcantggaa aggagganga ngcanggggtg 600  
 gatggggg 608

<210> 95  
 <211> 706  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(706)  
 <223> n = A,T,C or G

<400> 95  
 agaataatga gtacacaagt ttttgcctat ccctttcctt aaaatataag aaataaagga 60  
 gctcttacga aaaacccagg ttagaccgca taaaaataa agtgaacagt gaggtggtag 120  
 caagacttct ttttagaaaa gaaagcattt acctgcctgt ctgtaagggt gaaatttcat 180  
 cagtttgcaa acgataagaa atgcagactt gctcttgata gaaatgctta gaaacactct 240  
 ggggggggaa aaagctcctt ccataactg tgagacattt gttaagtga atctattgtt 300  
 tatcagcttt taaggataaa aaaggtattt taaaagttgg atatttagga tatttgagga 360  
 tattccttta tgagctctcc atatccttct tgagaaactg gttaaaaaag gaataggggt 420  
 tgagtgttac agagagtagt ctgaagattc ctgtgtaaaa gcaaagctaa caagcaatga 480  
 agacatgaag caaaatacta atctaattgt gtaaagaaag gatattttta taagttcctt 540  
 ctgcttgctg ctaagagttt gctaaagtgt catgaattat tctggttatt actaaagttt 600  
 ctatgaaaca agtagatttt aagaataaat gtttctggaa aagaactatg ttatgatttt 660  
 gtanaaatgt aaagattact tgaggtgttt aaaataaatt tttcat 706

<210> 96  
 <211> 719  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(719)  
 <223> n = A,T,C or G

<400> 96  
 ccagggcctg tccctgcccc agtggcggct gcgctgggtt caggcccagg gggccctgna 60  
 aaagctgtgc agctgaagag agggttcaaa cgaagccga gaacttgaca ctgttcaccc 120

```

caacacctca cctccccag gacatttga agaaagcagc gccaggattc ctcggcagtc 180
gtccccaccc gcacctgcag tccccctatg tgctgttctg ctgccccact cagctcctgg 240
acctgtcctt ttcattccgc taaagcacc cctaaaaccc ctctatcact ttcattctca 300
gcaaaaagta attgagcacc tcctctaggc gctggggagt ccacactgaa caaaagaaac 360
agaaaaccct gtcttccagc agttgagttc tagggcaggg agacagagtt tacaagataa 420
ggaaaatata tatgtagtat gctgcaagtt aactgctgtg tgggaaatcc agcaggggggt 480
gggatgtgtg atttgaattg agggccacac tgcccagggtc gtgctccgtc aaggggtgag 540
caggagcaac aggggtggct gagtaanggc ttgcagctgg aggcaacagc acatgcaaaag 600
gccctgagcc aggatgtgct gcaataangg ccactgaggg ggacagtgtg ngttgggggg 660
tgangaatgt attaaaaaga ngaaaaattg ccttttaaaa aaaaaaaaaa aaaaaaaaaa 719

```

```

<210> 97
<211> 572
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (1)... (572)
<223> n = A,T,C or G

```

```

<400> 97
gcttggctaa tttccaaatt attgcataat atgttctacc ttaagaaaac aggtttatgt 60
aacaaagtaa tgggtgtgaa tggatgatgt cagttcatgg gccttttagca tagttttaag 120
catccttttt tttttttttt ttttgaaagn gtgttagcat cttgttactc aaaggataag 180
acagacaata atacttcact gaatattaat aatctttact agttttacctc ctctgctctt 240
tgccaccoga taactggata tcttttcctt caaaggaccc taaactgatt gaaatttaag 300
atatgtatca aaaacattat ttcatttaat gcacatctgt tttgctgttt ttgagcagtg 360
tgcagtttag ggttcattga aaatcattga accacatgtg taacaactga atgccaaatc 420
ttaaactcat tanaaaaaata acaaattagg ttttgacacc cattcttaat tggaataatg 480
gatcaaaaat agnggntcat gaccttacca aacacccttg ctactaataa aatcaaatat 540
cccttagaag ggtatgtatt tttagttagg gg 572

```

```

<210> 98
<211> 520
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (1)... (520)
<223> n = A,T,C or G

```

```

<400> 98
tcagcctggg gtcctccact tctggaggcc tctaggctct atcagcaact gggggacana 60
acagcagagc tggagagtct ggagctgcta gttgaggcct tgaatgtccc atgcagttcc 120
aaagccccgc agtttctcat tgaggtagaa ttactactgc caccacctga cctagcctca 180
ccccttcatt gtggcactca gagccagacc aagcacatac tagcaagcag ggcaggagac 240
gctgcagagc attacttga cctgctggcc ctggtgctgg atagctcgga gccaaaggttc 300
tccccacccc cctccccctc agggccctgt atgcctgagg tgtttttga ggcagcggta 360
gactgatcc aggcaggcag agcccaagat gccttgactc tatgtgagga gttgctcanc 420
cgcacatcat ctctgtacc caagatgtcc cggctgtggn aagatgccan aaaangaacc 480
aangaactgc catactgncc actctgggtc tctgccaccc 520

```

```

<210> 99

```

<211> 470  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(470)  
 <223> n = A,T,C or G

<400> 99  
 tttttttcaa agcaattgtg acacctacct gtggaccaag gaaaaaacca aatcatccnc 60  
 aatctttcag ttccacaaca acaccattat aaaaacgagt tgatctgaag tggttccaaa 120  
 tctttcttat cttcaagatc tatgattagt aattctgact cgttgcaaag caacatat 180  
 tttgagatac tgttgtgagc cccgggagaa tgtcagagtc ctgtccactt ggacatgggt 240  
 ggagggggagg ttgggcaggg ctgataaaaag actagtaacg ctcgtgtctc ttcagtctta 300  
 ttttgcaaac atcatctata aagggttttg ctactccaag ttttggtagc ccaagctcat 360  
 caaagcatgt gtctattatg tgnctancat agtaaaaatg gctctaaatt gcataataat 420  
 gccagattt taataatcta ttgnttcana agaaaacaac agnannngnt 470

<210> 100  
 <211> 570  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(570)  
 <223> n = A,T,C or G

<400> 100  
 agtttttatt tttttctgta acgcccattc tgcctgcagt agaccctttg ggcttaagan 60  
 gtcttacaga ttctaatagt ttccctcttg tagatgtgtt aacatattta aaatacttag 120  
 gacatggcct ggtacgtgac agatgggtgt taacatcatt gtcatatgaa gaacactctc 180  
 tgtgggcca aaggaggtgt ccttaaaaatt cttcatcttg cctgattttg tttgcataac 240  
 ttctggagag tctgtgtcct cctcatctag gccaccttcc cattttttgt gaaagatgac 300  
 cttgtccctg cagtctccgc caggagtcac acagcttttc acagcagctg ccatctttaa 360  
 ttctttttcc tgagattcca gcaagggtgt gacattgtca cttttttgtt ctagactctt 420  
 ttaaattttc tgcatttgcc tgaaaagcac ccctgtaaga atagatttct catgggtcta 480  
 aaaattattc ccaagaatac cttacttggt tcaaaaagcag actgtttctc ttcatttcat 540  
 ctcaaatcan acttctgggc aagatgttct 570

<210> 101  
 <211> 365  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(365)  
 <223> n = A,T,C or G

<400> 101  
 atcaataaac caaggtaatg cctnagggtt ccctcccagt cctcncatca gctctggcct 60  
 catcaccaag gtcacanagn acacagggga gggggaanac ccacncacac tccttggaat 120  
 gggtcctgtt atttatgctt ggngcncagn catattanaa gaaaaaaaaa agctttgtat 180

```

tntttttccn catatnatgg ntgctgttta cacaccctgc caatgcntta gcnntggaga 240
gcttttttgca atatgnnggg gaaaggggag ggaggggaatg aaagtgccaa agaaaacatg 300
ttttaanaaa ctnggggtttt atacaataga atgttttcta gcagaaaaaa aaaaaaaaaa 360
aaaaaa 365

```

&lt;210&gt; 102

&lt;211&gt; 546

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1) ... (546)

&lt;223&gt; n = A,T,C or G

&lt;400&gt; 102

```

antgacacac tagccccctg cgcctcggc atgaacagca tgggcacccat ggatggcacg 60
ctgatgggtgg taggaagcan aagggtcctt gggcagaaaag ggcttggtnt gaggtgaaac 120
ccnccttca agcctgggat ggcctgagggc ctggggggccg ggctgccagc tntgggtagt 180
ntgtgggtgg agtganaatt tatggtgctt tttccgggccc tgctcatgga ccaatcagca 240
tgcacttccct ccttntgag ccataaaaaa ccctggattc agccagactt ggacanatgt 300
caatactacc aactgtggga agaagctacc cacttcagga ctcccttgact cctcaggatg 360
acctgcctac anaaagaagc taccactat gggtnntntn tntgctgana gctggatact 420
ngtcaggatg acctgcctgc anaaanagct acccnccatg ggtntcctnt gaactgttct 480
gttgcccaat aaagctcctn tcctttttgc tcnccttcca aaaaaaaaaa aaaaaaaaaa 540
aaaaaa 546

```

&lt;210&gt; 103

&lt;211&gt; 376

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1) ... (376)

&lt;223&gt; n = A,T,C or G

&lt;400&gt; 103

```

ctcaactccc ncttcacctt atgtaatgtc tgggcctgag atttctctct ggctttantt 60
tcttgatgtt cctgtgatat ggcttctgcc agcaatgaaa acaaggctcct tatagtagtn 120
ctccatggag tccagttcat cctgggtggcg tctcctttgt tcatctcgct tttctttggc 180
atagtttctt aggtctcgta atctttgctt ttgaatgttt aaaccttctt caaacagttt 240
cttaaatatc atttcttccc gggctctcat tctcatcatt tttgcacaca actgaactct 300
ataatcatca taatattttc gagcacgaac gatttgctgn cgatnttctt ttatctttga 360
ttgggncaac ctttgc 376

```

&lt;210&gt; 104

&lt;211&gt; 700

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1) ... (700)

&lt;223&gt; n = A,T,C or G



&lt;400&gt; 104

tttgaaaata	cagattttta	ccaacttttg	attctttttt	agttatatgt	ttgtctttcc	60
tttttaaatt	gttcaaaact	atTTTTta	ggtcaagtta	ctaacacttg	aaaatcagat	120
actgcaccaa	atacagtgtt	tttccgtagt	gtttttaatg	agtgcaccta	ttactactgt	180
gcgagaattc	atgtttttacc	agtcattgtt	atattacaaa	cagacttgca	tgattaacca	240
gttggttacac	ttactttttc	aagttggagt	atatatgact	cagtgcagac	tggtctctct	300
tatgtgaatg	cacacatgca	gaaatgcaga	gtcaatttta	catgcccata	aagacatttg	360
taaagaattc	agctcttatg	gtctgttgta	taaagtgtga	tctaggcact	ttggaatttg	420
acctcacaga	tgttacaact	tgatcagtcg	tttgacctaa	tttgtggnag	ctatctgtat	480
gttttgcaat	cttaatacag	acatgctttc	caaaaagatt	aatacagaac	catcctgccg	540
ttttggataa	gtctatccag	ctgtggaaag	ggcaacctgt	ggtttctctg	tactgggtgt	600
taatggggga	agaatatgaa	cagcttttaa	gagctgtgta	ttgggggtac	tactattaaa	660
aaataagatc	tgacgagtc	tgactggcct	ttgggtggcc			700

&lt;210&gt; 105

&lt;211&gt; 729

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1)...(729)

&lt;223&gt; n = A,T,C or G

&lt;400&gt; 105

taaacccttc	acactgagaa	ctgcttgtgt	gggagagagc	tggttgggtt	gatcttttcc	60
gagtgtgact	tacctcttc	aaggggatgt	ttaagcttct	cgggcagaag	tggtgtgtct	120
attcctgaca	ccaaacaccg	tggtatatgt	ggttgtcaca	ctcagctagt	gatgataaag	180
gtgttcttaa	atatgttagc	tttcagtttt	cctgaggaag	caattttatg	gatacttccc	240
cctccctctc	aagtgaggaa	tagcagagca	aattttattt	ggaacttaaa	ccaatagtta	300
taaccaatag	tttcaacctc	ctgcctcacc	actgcttcc	tcctgagctc	tttccccaca	360
cctcaaaaag	agtacaaaag	gattccatct	gcagaggtaa	attctttggt	taaaaaagta	420
ctgtttttct	tatcttttct	ggttctccta	ggtatcagaa	caaggtttaa	tagaaaatcct	480
taaaaaagta	agccaacaaa	cagaaaagac	aacaacagt	aaagtaagt	tccccanatg	540
ctgttgccaa	atgaaaagat	ggatacttta	aagattaatg	ttgagtatac	atctaccaca	600
catatttttc	agcccanaga	cattttcctt	ttgtcaaaaa	cgtgaaagtt	tgggggagaaa	660
ngctgaatct	gttgggggag	ggttctaatt	tttataggct	cttgactcca	ttccccctt	720
ttaattcac						729

&lt;210&gt; 106

&lt;211&gt; 481

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 106

gaaaagcaga	gtcagctact	gagagctctg	gaaagaagga	tgtcaagaag	gtgaaatcct	60
aaagcctaga	aataaagttt	taaattggaa	actgctattt	tcttgttccc	atcttcaaat	120
gctaattgcc	agttccagt	tattcatggt	actctaagaa	aaatctcttt	ggttttgatt	180
tcttgcatat	tttatatatt	ttacaatgct	tctacactga	aatgtgtagc	tttatatttt	240
atggcattct	agtatttttg	tgtactgtat	tttgtgcatt	tcatgtcttc	atcaaaatcc	300
tctcagtcct	tggtcttttg	aagcttgtgc	tgagggttta	gcttttctat	gttttatatg	360
ccgctgcttt	gaaagagaac	ctagattcta	tagttgtatt	attgttgttt	catactttaa	420
atttatatgg	ctgtggaaaa	acgaattaaa	atgttttgag	gagaaaaaaa	aaaaaaaaaa	480
a						481

<210> 107  
 <211> 519  
 <212> DNA  
 <213> Homo sapien  
 <220>  
 <221> misc\_feature  
 <222> (1)...(519)  
 <223> n = A,T,C or G

<400> 107  
 cagctcctct cctgccagag ctaggcaggc gccgaagtag ccgcatggcc ccgtcagant 60  
 accccagggg ctggagagcc aacctcaaag gcaccatccg tgagacagge ctggagacca 120  
 gctccggtgg gaagctggct ggccatcaga agaccgtccc cacggctcac ctgacttttg 180  
 ttattgactg caccacggg aagcagctct ccctggcagc aaccgcatca ccacccaag 240  
 cccccagtcc caatcgaggg cttgtcaccc caccaatgaa gacctacatc gtgttctgtg 300  
 gggaaaactg gcccactctg actcgggtga ccccatggg tgggggatgc cttgccagg 360  
 ccagggccac cctgccgctc tgcagagggg ctgtggcctc agcttccttc ccagtcagcc 420  
 cgctctgccc ccaggagggt cccgaggcta aggggaaacc cgtgaaggct gcgcctgtga 480  
 ggtcttcaac ttggggaaca gtcaaggact cactgaaag 519

<210> 108  
 <211> 669  
 <212> DNA  
 <213> Homo sapien

<400> 108  
 ggacaatgaa gactgaacta tcgcacatta cctaagaaag atgggaattg acatgcacat 60  
 cacaattgta tacacaacag aaattattga atcatgagat atacattccg gtgtgtgaca 120  
 gattggcaca tgacataatc tgggttcttt atagactcag ttgttttggg gcgatctaga 180  
 ttatcaagag aagagccttc atgcttagct ttatcttgta ccaaccacca gggcccttgc 240  
 ttctgagcag gaagcagctg gggaatagge tctttctctt aatgacttcc aacatagtgc 300  
 tctcaaactt tactcctcca gaaggccacc ctacctggc tatggctact tcagaaaaaa 360  
 cttggcctct ggtataatag agcagaatca tcacctcaca ttctatttca agccaaagtc 420  
 aatatctcaa aggctgggtc tgtgatttat ttggctcttg ggagctccta ctgaaagtgc 480  
 tgaaatgtcg tactgacact tcagacttat agctacctag actccaagta agatttatct 540  
 ctgactggag ggtttctcct attaaaaacc aaagagtgtg ggggtgccttc acctgctagg 600  
 taatcttcta tgccctaatt ggaagaatgg gagcagcaga caagtaagtg caggaaggag 660  
 aaccaaaagc 669

<210> 109  
 <211> 349  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1)...(349)  
 <223> n = A,T,C or G

<400> 109  
 tttcaagccg gatatttggc ctgcttaaac cacttaaatg tanttaatga cagatggttt 60  
 gaggtttaaa agtcttctgg agaaagcccc ccagagaaca ttccctttga agccccatgt 120  
 aaaaatacgt gtgggagaga aagtgttttc tctgacttct gctgacagtg gctaaaaactc 180

tgaactgtca ggagtattca aaataagact gccttgtagg taagcctgtg gtagcttttt 240  
 tgagcacagg ataaaatact tgagtctttg cttaaagtgt actttctcaa tgaggctttg 300  
 tatgactaaa taaaatctgt ataatcccca aaaaaaaaaa aaaaaaaaaa 349

<210> 110  
 <211> 337  
 <212> DNA  
 <213> Homo sapien

<400> 110  
 ggcctttccc actggtccat ctggttttct ctccagggtc ttgcaaaatt cctgacgaga 60  
 taagcagtta tgtgacctca cgtgcaaagc caccaacagc cactcagaaa agacgcacca 120  
 gccagaagt gcagaactgc agtcactgca cgttttctac tctagggacc agaaccaaac 180  
 ccaccttttc tacttccaag acttattttc acatgtgggg aggttaatct aggaatgact 240  
 cgtttaaggc ctattttcat gatttctttg tagcatttgg tgcttgacgt attattgtcc 300  
 ttgtattcca aataatatgt ttccttcctt caaaaaa 337

<210> 111  
 <211> 552  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1) ... (552)  
 <223> n = A,T,C or G

<400> 111  
 attttgtcaa gtttctttaa tggctgaaca gaaagaagct tcaagtaatg gagaaggcat 60  
 tgtctgagtg cagctgcttt cctggacgcc tgtgccgttc ctgtcttcca aatcctatgc 120  
 ttggaggccc ctggagggtac atttttgcca ggaaccaacc tgaccttaaa aagatgagtg 180  
 tgacacagcc ggctgggagc gaggatggag gtgccacagg acaccacctg cccacgccct 240  
 ggccagcctg gccatgctgc cgagtgcggg gagggcaccc caccagagg gcacagggca 300  
 aaccttaagc acgggggtat tgcccttgaa gccccagggg atgccctgtg ccggatcctc 360  
 atgcctcatt gactagcctg cttgctgaag gagccccagg gcctgagcct gcaacactgc 420  
 aaggggtgag aagggcatgc tgcgtggggc gccactggac tcaaacctca cattanaagc 480  
 taaaaagaac cccaaatgcg cttcaanagc cccaccaacc cccaagccan gctcatncct 540  
 gacacantga at 552

<210> 112  
 <211> 115  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (1) ... (115)  
 <223> n = A,T,C or G

<400> 112  
 gtcctaggta canaaacttc ccaaacaana actatgencg ancnetaccc ctteagagtgc 60  
 tctncnttct gtcctgggtc ccatggnttg ccccatatta taaggtagc tggga 115

<210> 113  
 <211> 649

<212> DNA  
<213> Homo sapien

<220>  
<221> misc\_feature  
<222> (1)...(649)  
<223> n = A,T,C or G

<400> 113  
gcctgattaa aaactaagca gaagtagttt taacanaaat actcatgaaa atgtttngaa 60  
actgaaatgt aaacaactgt aatattaagg aaaccagaat caataaatca ctgtcttgcc 120  
agcacagcta cagagtaaca tgattcaggg gaggaaagtt ccttagagtt acttttataa 180  
ttcttttttt ttctctctta ggtttananaa tcttacaat ttaaacttta tccttttaaa 240  
attatttgaa cataatttag atattgtaag cttaaaatac aaatgtttat agataacctc 300  
tttaccataa actaatccct ggcaagccat ggctctcttt tttttttggg gtttaaagcc 360  
tgtaaacagt ttttctgaat gatcatgaac ttttcttggg tttagcacta ggatttagct 420  
atgaagagag ctcataggct ttcagggtgct aattgagatc tgccctgtta gaggcttggg 480  
gtgctagatt ggtcacattg acaccagtgg caggggaaggc atctatgagt ttgatgcttt 540  
ttatcacaca cttcagntgt ttagaaagtt attaccaata cttttaaaca acactccaag 600  
aaaatttgct atatttcttt ctcatcacta cagagagagt agatttccc 649

<210> 114  
<211> 650  
<212> DNA  
<213> Homo sapien

<220>  
<221> misc\_feature  
<222> (1)...(650)  
<223> n = A,T,C or G

<400> 114  
tggcgattgg tgttggcggg ctggctcagc tgggcagggg gtaactttac tgatttgggg 60  
gtgggttttta gtttaatttt tcttttctag cttcccatcg acggctcagt cgacagtgtg 120  
aatcagctga ggccatgtca ggagacggag ccacggagca ggcagctgag tatgtcccag 180  
agaagggtgaa gaaagcggaa aagaaattag aagagaatcc atatgacctt gatgcttggg 240  
gcattctcat tcgagaggca caggtttagt gatataggat tacatttcct tctctatggg 300  
tccaatcaca ctacttggtt ctgcagtga taatattttc ataatcctaa cattgtaaat 360  
gctgtttatt ggttttcaat tttagaatca acctatagac aaagcacgga agacttatga 420  
acgccttgtt gcccagttcc ccagttctgg cagattctgg aaactgtaca ttgaagcaga 480  
ggttactatt ttattttatt ttttcttata tcagtatttg cagcattcac tgtagtgata 540  
gaaaacaaag ttangaacat agccaattan gacaaggagg atttaaatgt gtcttacctt 600  
tattttgtaa aataggtata aaggagtaat taaaatgaat ttttgaaatt 650

<210> 115  
<211> 403  
<212> DNA  
<213> Homo sapien

<220>  
<221> misc\_feature  
<222> (1)...(403)  
<223> n = A,T,C or G

<400> 115

```

gtattatgac tttaaaaacc ccattattga aaagtacctg acaaggcagc tcacgaancc      60
caggcctgtg atcctggacc cggcggaccc tacnngaacc ttgggtggng gagacccaan      120
gggnnggagg canctggcac angaggctga ggcctgnctg aattacccat gctttaagaa      180
tngggatggg nccccantga gctcctgnat tctgctgggtg anacctcctg cttcctccct      240
gccattcatc cctgcccctc tccatgaagc ttganacata tanctggaga ccattctttc      300
caaanaactt acctnttgcc aaaggccatt tatattcata tagtgacang ctgtgctcca      360
tattttacag tcattttggg cacaatcgag ggttnctgga att                        403

```

```

<210> 116
<211> 397
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1) ... (397)
<223> n = A,T,C or G

```

```

<400> 116
cgaaagaaaa aaagtgatat tggaccctgg aaagattttg aaacttgagt ggtttgataa      60
cccttctatg tattgtaggg agaaaaaaaa aagtttattt tattccactg tcctccctta      120
aaagcatcat ttgancaata aatgaatatt gtctttaaac caagggttag ggaattttcc      180
tctctctctc tctctcctct ctctttctgt tcaaagaact tcaaacattt gggaccacct      240
ggtattctgt attttcactg gccatattgg aagcagttct agttgcattg tattgagttg      300
tgctggcagt agtttccatg cctgtcaatg tatcatagtc ctttgttgcc cagataaata      360
aatatttgat acgcttttaa aaaaaaaaaa aaaaaaaa      397

```

```

<210> 117
<211> 59
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1) ... (59)
<223> n = A,T,C or G

```

```

<400> 117
cacttatggg gacaatggga agactcttct tttncactgg actgtacctg gacntnnaa      59

```

```

<210> 118
<211> 751
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1) ... (751)
<223> n = A,T,C or G

```

```

<400> 118
catcaacata cccgctttat tgctgactca tgacaactaa tgggaagaca tggctcagat      60
gtgcagccac agtgagcttc tgaacatttc ttctcagact aagctcttac acacagtgtg      120
agttgaaaga aagaattgct tgacatggcc acaggagcag gcagcttcct gcagacatga      180
cagtcaacgc aaactcatgt cactgtgggc agacacatgt ttgcaaagag actcagagcc      240

```

```

aaacaagcac actcaatgtg ctttgcccaa atttaccat taggtaaatc ttccctctc
ccaagaagaa agtggagaga gcatgagtc tcacatggaa acttgaagtc agggaaatga
aggctcacca attatttgtg catgggttta agttttcctt gaaattaagt tcaggtttgt
ctttgtgtgt accaattaat gacaagaggt tagatagaag tatgctagat ggcaaagaga
aatatgtttt gtgtcttcaa ttttgctaaa aataaccag aacatggata attcatttat
taattgattt tggtaagcca agtcctattt ggagaaaatt aatagttttt ctaaaaaaga
attttctcaa tatcacctgg cttgataaca tttttctcct tcgagttcct ttttctggag
tttaacaaac ttgttcttta caaatagatt atattgacta cctctcactg atgntatgat
attaagttct attgcttact ttggatttct a

```

300  
360  
420  
480  
540  
600  
660  
720  
751

<210> 119  
<211> 591  
<212> DNA  
<213> Homo sapien

<400> 119

```

aggcttcttg tcacactgaa cacatccagc cacaggcacc agctgggttg gaccagcagc
cccagcatc ctcttgcaact ggctggcaca aaaagaaacc tgctgtatac ccccaaagt
gtccctttcc caattacctc tggggctctc tgctgcttgc ctctgctgct ctggactggg
agagcttctg tctgtgctg catgggtatt tagactgtgg gggagatgcc ccttcttata
gcaatggagg aggaaaacaa attcttgctc cctcagaat gagagtggct ctttctgatt
tgcaagggca ctatggtcag ggcaaaggca tggcccaggt gtttaagtac aggtgacgt
gtgcttatgc aatgggggtg taaggcaggc acgaagagtc caaaaaatct aggtggcctc
tcagctctgc cacctctagc tgcatgacct tgggcaagct atgtaacccc aattgcctgc
tccattaaag actgtgaagg tagaatgtt gtaaagctct taacagtatg taagccttca
ataaatttca gttttccctt tgttttcttg atcaaaaaaa aaaaaaaaaa a

```

60  
120  
180  
240  
300  
360  
420  
480  
540  
591

<210> 120  
<211> 652  
<212> DNA  
<213> Homo sapien  
  
<220>  
<221> misc\_feature  
<222> (1)... (652)  
<223> n = A,T,C or G

<400> 120

```

atattgcagt acattaaaac tgaggccag agatgtgatt tgcttgaggc cacacagcta
gatttttggg ggaagtgggc cttgaacaca gtgtactttc tgcagtttct gactgtaaaa
cccagtgtct gctctctgag ttccatttcc aagccccctt ccatcttgga cctatgtggg
ctccaccata ttcacacacc accgccacca cttgccaatg cctctcttaa agcaatatac
ccattcgttc tcttattggg aactggatgg atgaagcccc aaattcagcc ccaccacag
agaagccttc ctacactcag cctctgtoca cccttggaat atctttcaag ctctctcctc
caggaaagtg gggccccaac tcagtcactc caccaccttc caggctccctg aggtggttc
tactgtatcc ccatcacctc cacaactcca ctcacctctg acggctccat ccacctcacc
agtttgaagg cttgtggttt cagagaggag caatgctggg cagcgtgcc cagactccag
tgtttacaga tcaccagcat ttacaaccaa tccaatggcc agaagcctcc tctaaccana
aggagttctg aaggggcaga tgggggtgtg agtagtcggg gagtcgggat tg

```

60  
120  
180  
240  
300  
360  
420  
480  
540  
600  
652

<210> 121  
<211> 407  
<212> DNA  
<213> Homo sapien

```

<400> 121
ctctttagta tgacactggc aactgacggc actgcggata tgtttggaat aagggttcaaa    60
agaagaggct gttctaagaa agactaccga gtaatcaatc acctcctcag agaaagtctg    120
ctcaggaaac tctcctccca ccagccctg cacttggttc cttgcgttgc ctctgggtgg    180
aagctgttgt cccaacttc agaaagtgtc tgcagatttc acaggctctc tcttgggtca    240
agggaccagc tctgtgaaca cggcaagtaa caacagagag aaaacgtgag gagaaaagag    300
agaccgggat tttgaaatcc tgtttcctga atgccatctc accaggcaca cgaaagaata    360
aaattaagga atcaagaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaa                407

```

```

<210> 122
<211> 752
<212> DNA
<213> Homo sapien

```

```

<400> 122
caaatccctc ctatccaggc ttttctgtct ctaatacccc aagcgttacc cctgctcttc    60
cctcattccc ggggctgcag gcgcctctta cagtgcgagc tgtcacacca ctacctgtgg    120
ctgccacagc cccatcccca gctccagtcc tcccaggatt cgcctcagca ttcagttcca    180
atttcaactc cgctcttggt gcacaagccg gtttatcctc tggacttcaa gctgcaggca    240
gttctgtttt tccaggcctt ttgtccctcc cgggtatccc tgggtttcct cagaatcctt    300
cacaatcctc cttgcaagaa ttacagcata atgcggctgc acagtcagca ttgttacagc    360
aggtccattc agcttcggct ctggaagct atccagctca gcctgatggg tttcctagtt    420
atccttcagc gccaggaaaca ccattttctt tgcaaccaag cctgtcccag agtgggtggc    480
agtgaatact tttaactttt attctccttc agagcaacat cagaattgcc tgagaactgc    540
aatgaacaat ctgacaaatg tgaagctggc caaaagtcgg aaaatgagaa tgagggtaat    600
cctggagaaa ttgtgacaac aatttgaaaa ttgtggttgc attttaaaagt gtgaacactc    660
ccctatgtaa atatgctgac aataaattgg atggagaatg gtatttaaaa agtgtttgga    720
gacttttcac ctgtcctata aaaatttgaa tt                                752

```

```

<210> 123
<211> 401
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (1)...(401)
<223> n = A,T,C or G

```

```

<400> 123
atgaacaaaa tagaaataag catgcaacat gaacagctgg aagagagttt tcaggaacta    60
gtggaagatt accggcgtgt tattgaacga cttgctcaag agtaaagatt atactgctct    120
gtacaggaag cttgcaaatt ttctgtacaa tgtgctgtga aaaatctgat gactttaatt    180
ttaaaatctt gtgacatttt gcttatacta aaagttatct atctttagtt gaatattttc    240
ttttggagag attgtatatt ttaaaatact gtttagagtt tatgagcata tattgcattt    300
aaagaaagat aaagcttctg aaatactact gcaattgctt cccttcttaa acagtataat    360
aatgcttag ttgtgatnaa aaaaaaaaaa aaaaaaaaaa a                                401

```

```

<210> 124
<211> 103
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature

```

&lt;222&gt; (1)...(103)

&lt;223&gt; n = A,T,C or G

&lt;400&gt; 124

aagataactt gctgggacca catgcctgat tgccacagct gtcacacgt tatttaatga 60  
 acctagtgtc gaagacagtg aaaagggtcc attgacnctg gcn 103

&lt;210&gt; 125

&lt;211&gt; 1024

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1)...(1024)

&lt;223&gt; n = A,T,C or G

&lt;400&gt; 125

gagggcagtg aggagcgagg agcgggcaga ggcagctccg gcggccgaga ggagggagcg 60  
 cggcgagag aggagggggt tgcgccccgt agaaatgtca atcagagcct ggacccttgc 120  
 gctcccgcac cagccccgc ctccgccctg cagaagccca agatctgtgc cctcgcgag 180  
 actgccacaa gcccgacaa cccgcgccgc tcgctcccg gcgcgggggg gtctccaccg 240  
 ggggcagcgg tcgcgccttc cgccctgcag ctctctccgg ccgcgcgcgc cgccgcgcgt 300  
 cacagactgg tctcagcgcc gctgggcaag ttcccggctt ggaccaaccg gccgtttcca 360  
 ggcccaccgc ccggccccgc ccgcaccgc ctctccctgc tgggctctgc ccctccgcac 420  
 ctgctgggac ttcccgagc cgcgggccac ccggtgcgc cgccgcctt cgctcgccca 480  
 ggggagcccg aaggcggaac agatcgctgt agtgccttgg aagtggagaa aaagttaactc 540  
 aagacagctt tccagcccg gcccaggcgg ccccgaaacc atctggagcg cgccctggtc 600  
 ttatcggtc tctcctcctc ctagtctctt aaaaaaaaca aaaaaacaaa aaaaactttt 660  
 ttaaatcggt gtaataattg tataaaaaaa atcgctctgt atagttacaa cttgtaagca 720  
 tgtccgtgta taaataccta aaagcaaaac taaacaaaga aagtaagaaa aagaaataaa 780  
 accagtctc ctcagccctc cccaagtgc ttctgtggca ccccgcatte gctgtgaggt 840  
 ttgtttgtcc ggttgatttt ggggggtgga gtttcagtga gaataaacgt gtctgccttt 900  
 gtgtgtgtgt atatatacag agaaatgtac atatgtgtga accaaattgt acgagaaagt 960  
 atctattttt ggctaaataa atgagctgcc tgccactttg nctataaaaa aaaaaaaaaa 1020  
 aaaa 1024

&lt;210&gt; 126

&lt;211&gt; 214

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 126

Arg Pro Arg Ile Arg His Glu Glu Gly Ser Glu Glu Arg Gly Ala Gly  
 1 5 10 15  
 Arg Gly Ser Ser Gly Gly Arg Glu Glu Gly Ala Arg Arg Arg Glu Glu  
 20 25 30  
 Gly Leu Ala Pro Arg Arg Asn Val Asn Gln Ser Leu Asp Pro Cys Ala  
 35 40 45  
 Pro Ala Pro Ala Pro Ala Ser Ala Leu Gln Lys Pro Lys Ile Trp Ser  
 50 55 60  
 Leu Ala Glu Thr Ala Thr Ser Pro Asp Asn Pro Arg Arg Ser Pro Pro  
 65 70 75 80  
 Gly Ala Gly Gly Ser Pro Pro Gly Ala Ala Val Ala Pro Ser Ala Leu  
 85 90 95



Gln Leu Ser Pro Ala Ala Ala Ala Ala Ala His Arg Leu Val Ser  
 100 105 110  
 Ala Pro Leu Gly Lys Phe Pro Ala Trp Thr Asn Arg Pro Phe Pro Gly  
 115 120 125  
 Pro Pro Pro Gly Pro Arg Pro His Pro Leu Ser Leu Leu Gly Ser Ala  
 130 135 140  
 Pro Pro His Leu Leu Gly Leu Pro Gly Ala Ala Gly His Pro Ala Ala  
 145 150 155 160  
 Ala Ala Ala Phe Ala Arg Pro Ala Glu Pro Glu Gly Gly Thr Asp Arg  
 165 170 175  
 Cys Ser Ala Leu Glu Val Glu Lys Lys Leu Leu Lys Thr Ala Phe Gln  
 180 185 190  
 Pro Val Pro Arg Arg Pro Gln Asn His Leu Asp Ala Ala Leu Val Leu  
 195 200 205  
 Ser Ala Leu Ser Ser Ser  
 210

<210> 127  
 <211> 507  
 <212> PRT  
 <213> Homo sapien

<400> 127  
 Met Ser Phe Pro Gln Leu Gly Tyr Gln Tyr Ile Arg Pro Leu Tyr Pro  
 1 5 10 15  
 Pro Glu Arg Pro Gly Ala Ala Gly Gly Gly Gly Gly Ser Ser Ala  
 20 25 30  
 Gly Gly Arg Ser Gly Pro Gly Ala Gly Ala Ser Glu Leu Ala Ala Ser  
 35 40 45  
 Gly Ser Leu Ser Asn Val Leu Ser Ser Val Tyr Gly Ala Pro Tyr Ala  
 50 55 60  
 Ala Ala Ala Ala Ala Ala Ala Ala Ala Gln Gly Tyr Gly Ala Phe Leu  
 65 70 75 80  
 Pro Tyr Ala Thr Glu Leu Pro Ile Phe Pro Gln Leu Gly Ala Gln Tyr  
 85 90 95  
 Glu Leu Lys Asp Ser Pro Gly Val Gln His Pro Ala Thr Ala Ala Ala  
 100 105 110  
 Phe Pro His Pro His Pro Ala Phe Tyr Pro Tyr Gly Gln Tyr Gln Phe  
 115 120 125  
 Gly Asp Pro Ser Arg Pro Lys Asn Ala Thr Arg Glu Ser Thr Ser Thr  
 130 135 140  
 Leu Lys Ala Trp Leu Asn Glu His Arg Lys Asn Pro Tyr Pro Thr Lys  
 145 150 155 160  
 Gly Glu Lys Ile Met Leu Ala Ile Ile Thr Lys Met Thr Leu Thr Gln  
 165 170 175  
 Val Ser Thr Trp Phe Ala Asn Ala Arg Arg Arg Leu Lys Lys Glu Asn  
 180 185 190  
 Lys Met Thr Trp Ala Pro Arg Ser Arg Thr Asp Glu Glu Gly Asn Ala  
 195 200 205  
 Tyr Gly Ser Glu Arg Glu Glu Asp Glu Glu Glu Asp Glu Glu Glu  
 210 215 220  
 Ser Lys Arg Glu Leu Glu Met Glu Glu Glu Glu Leu Ala Gly Arg Gly  
 225 230 235 240  
 Gly Gly His Gly Gly Arg Gly Ala Gly Arg Arg Arg Asp Glu Glu

245 250 255  
 Ile Asp Leu Glu Asn Leu Asp Ser Ala Ala Ala Gly Ser Glu Leu Thr  
 260 265 270  
 Leu Ala Gly Ala Ala His Arg Asn Gly Asp Phe Gly Leu Gly Pro Ile  
 275 280 285  
 Ser Asp Cys Lys Thr Ser Asp Ser Asp Ser Ser Glu Gly Leu Glu  
 290 295 300  
 Asp Arg Pro Leu Ser Val Leu Ser Leu Ala Pro Pro Pro Pro Val  
 305 310 315 320  
 Ala Arg Ala Pro Ala Ser Pro Pro Ser Pro Pro Ser Ser Leu Asp Pro  
 325 330 335  
 Cys Ala Pro Ala Pro Ala Pro Ser Ser Ala Leu Gln Lys Pro Lys Ile  
 340 345 350  
 Trp Ser Leu Ala Glu Thr Ala Thr Ser Pro Asp Asn Pro Arg Arg Ser  
 355 360 365  
 Pro Pro Gly Ala Gly Gly Ser Pro Pro Gly Ala Ala Val Ala Pro Pro  
 370 375 380  
 Thr Leu Gln Leu Ser Pro Ala Ala Ala Ala Ala Ala Ala His  
 385 390 395 400  
 Arg Leu Val Ser Ala Pro Leu Gly Lys Phe Pro Ala Trp Thr Asn Arg  
 405 410 415  
 Pro Phe Pro Gly Pro Pro Ala Gly Pro Arg Pro His Pro Leu Ser Met  
 420 425 430  
 Leu Gly Ser Ala Pro Gln His Leu Leu Gly Leu Pro Gly Ala Ala Gly  
 435 440 445  
 His Pro Ala Ala Ala Ala Ala Ala Tyr Ala Arg Pro Ala Glu Pro Glu  
 450 455 460  
 Ser Gly Thr Asp Arg Cys Ser Ala Leu Glu Val Glu Lys Lys Leu Leu  
 465 470 475 480  
 Lys Thr Ala Glu Gln Pro Val Pro Arg Arg Pro Gln Met Arg Leu Asp  
 485 490 495  
 Ala Ala Leu Val Leu Ser Ala Leu Ser Ser Ser  
 500 505





*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 00/17536

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C07K14/47 C12N15/63 C12N5/10 C07K16/18  
C07K19/00 C12N15/62 A61K38/17 A61K48/00 A61K39/395  
C12N5/06 G01N33/574 G01N33/68 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K A61K G01N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EMBL, EPO-Internal, WPI Data, BIOSIS

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL 'Online! Entry AI681917, 27 May 1999 (1999-05-27) ROBERT STRAUSBERG: "tx50f03.x1 NCI_CGAP_Lu24 Homosapiens cDNA clone" Database accession no. AI681917 XP002160838 the whole document</p> <p style="text-align: center;">--- -/-</p>	<p>7-10, 58-60</p>

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*G\* document member of the same patent family

Date of the actual completion of the international search

20 February 2001

Date of mailing of the international search report

06.03.2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Montero Lopez, B

# INTERNATIONAL SEARCH REPORT

Int'l. Application No.

PCT/US 00/17536

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>           DATABASE SWALL 'Online!            Entry IRX3_Mouse,            15 July 1998 (1998-07-15)            Database accession no. P81067            XP002160839            the whole document            &amp; ANTJE BOSSE ET AL.: "Identification of            the vertebrate Iroquois homeobox gene            family with overlapping expression during            early development of the nervous system"            MECHANISMS OF DEVELOPMENT,            vol. 69, no. 1,2, December 1997 (1997-12),            pages 169-181, XP000982338         </p>	1
A	<p>           WO 99 14230 A (WASHINGTON UNIVERSITY)            25 March 1999 (1999-03-25)            the whole document         </p>	1,2,4-60

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US 00/17536

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
  
Although claims 21, 22, 29-31, 34, 37-39 and partially claim 35 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:  
  
3, and partially 1, 2, 4-60
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☒ No protest accompanied the payment of additional search fees.

## FURTHER INFORMATION CONTINUED FROM PCT/ISA 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: Partially 1, 2, 4-60

Polypeptide encoded by a polynucleotide of SEQ ID NO:1  
analogous thereof; polynucleotide of SEQ ID NO:1 and analogs  
thereof, expression vector and host cell comprising the  
same; antibody for the polypeptide and diagnostic kit  
comprising it; fusion protein comprising the polypeptide and  
polynucleotide encoding the same; pharmaceutical composition  
and vaccine comprising any of the above; use thereof for  
inhibiting the development of cancer; pharmaceutical  
composition and vaccine comprising an antigen-presenting  
cell expressing the polypeptide and use thereof for  
inhibiting the development of cancer and removing tumor  
cells from a biological sample; use of the above for  
stimulating and/or expanding T cells and T cell population  
so prepared; use of such T cells for inhibiting the  
development of cancer; use of the polypeptide and  
polynucleotide for determining the presence of cancer and  
monitoring the progression of cancer in a patient;  
oligonucleotides hybridizing to the polynucleotide and  
diagnostic kit comprising them.

2. Claims: Partially 1, 2, 4-60

Idem as subject 1 for SEQ ID NO:3

3. Claims: 3 and partially 1, 2, 4-60

Idem as subject 1 for polynucleotides of SEQ ID NOs:4 and  
125 and corresponding polypeptide of SEQ ID NO:126

~~4-56~~ Claims: Partially 1, 2, 4-60

Idem as subject 1 for respectively SEQ ID NOs:8, 9, 11, 12,  
14, 15, 17, 26, 30, 31, 35, 36, 47, 52, 55, 60, 62, 63, 66,  
69, 71, 72, 74-78, 81, 83-87, 92, 94, 97, 99-104, 107, 108,  
110, 112, 113, 117, 118, 120-122.

~~57~~ Claims: Partially 25-53

Vaccine comprising an antigen-presenting cell expressing a  
polypeptide encoded by polynucleotide of SEQ ID NO:2 or  
analogous thereof; use of such antigen-presenting cell for  
inhibiting the development of cancer; use of T cells with  
react with a polypeptide encoded by SEQ ID NO:2 or analogs  
thereof for removing tumor cells from a biological sample  
and its use for inhibiting the development of cancer; use of  
the above for stimulating and/or expanding T cells, T cell



## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

population so prepared and use thereof for inhibiting the development of cancer; use of the polypeptide encoded by SEQ ID NO:2 and corresponding polynucleotide for determining the presence of cancer and for monitoring the progression of cancer.

~~53-124~~ Claims: Partially 25-53

Idem as subject 5 for polynucleotides or respectively sequences SEQ ID NOs: 5-7, 10, 13, 16, 18-25, 27-29, 32-34, 37-46, 48-51, 53, 54, 56-59, 61, 64, 65, 67, 68, 70, 73, 80, 82, 88-91, 93, 95, 96, 98, 105, 106, 109, 111, 114-116, 119, 123, 124

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. Application No

PCT/US 00/17536

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9914230 A	25-03-1999	US 5922836 A	13-07-1999
		AU 9373798 A	05-04-1999
		BR 9812472 A	19-09-2000
		CN 1277614 T	20-12-2000
		EP 1037901 A	27-09-2000
		NO 20001358 A	12-05-2000